DOCTOR OF PHILOSOPHY

Assessment of marine recreational fisheries using: social media, fisheries dependent data and image analysis

Monkman, Graham

Award date:
2019

Awarding institution:
Bangor University

DOCTOR OF PHILOSOPHY

Assessment of marine recreational fisheries using: social media, fisheries dependent data and image analysis

Monkman, Graham

Award date:
2019

Assessment of marine recreational fisheries using: social media, fisheries dependent data and image analysis

A thesis presented to Bangor University for the degree of Doctor of Philosophy

By

Graham George Monkman

School of Ocean Sciences

Bangor University

September 2018:

THE FISHERIES SOCIETY OF THE BRITISH ISLES

Cefas

PRIFYSGOL BANGOR UNIVERSITY

a. Summary

Novel methods are required to improve knowledge on the activity of marine recreational fishers (which can be impactful); the stocks they prosecute, and the ecosystems the fishers and

their quarry interact with. “Traditional” survey methods rely on complex randomised designs based on sound statistical sampling methods and are the “gold standard” for evidence collection. But these methods are costly and logistically complex to deliver, which partially explains the relatively poor understanding of marine recreational fisher activity in the majority of recreational fisheries, even in developed countries. This thesis examines two separate (but interlinked) approaches to enhance knowledge acquisition. Firstly, two separate methods are described which use the local ecological knowledge of fishers to describe proxies for the estimation of the spatial and temporal distribution of effort. These proxies are validated against the best available ground truth directed-survey data. One method exploits social media, which can pose unfamiliar ethical questions to ethical research boards and the peers of researchers who propose to use social media. Consequently a review of the ethical issues surrounding the use of social media as a source of scientific data for fisheries research is provided. The second approach automatically derives accurate estimates of a morphological measurement (total length) of the European sea bass (*Dichentrachus labrax*) under real survey conditions (i.e. with limited control of the camera and related paraphernalia). There were two aspects to the length estimation process; (i) images were corrected for distortion, and length estimates corrected for parallax effects; and (ii) machine vision (transfer learning using three pretrained regional convolutional neural networks) and a machine recognisable marker were used to detect European sea bass in images taken with different cameras and on different angling platforms. These detections, together with the methods validated in (i) provided accurate length estimates with a percent mean bias error of -0.1%.

b.Contents

c. List of Tables

d.List of Figures

e. List of Equations

f. Acknowledgements

Participant	Contribution
Armstrong, Mike	For providing advice on the marine recreational fishing sector, on conducting recreational fishing surveys and the results of the Sea Angling 2012 survey.
Aron, Henry	For providing the FishMap Môn GIS layers.
CEFAS	For allowing Kieran Hyder to be my supervisor
Fisheries Society of the British Isles	Providing the PhD studentship
Hyder, Kieran	Supervisor
Jones, David	of the Pembrokeshire Coastal Forum for providing advice and the GIS layers for the Wales Activity Mapping project
Kaiser, Michel	Supervisor
Owen, Jason	Previous skipper of the charter boat the Haf Aled who allowed me aboard his charter vessel to take photographs of sea bass.
Richardson, Elizabeth	For providing advice on the data she collected as part of her 2006 PhD.
Vidal, Franck	Supervisor
Wm Morrison Supermarkets plc.	For allowing me to take pictures of sea bass in-store.

h.Abbreviations

Abbreviation	Definition
ACC	Autocorrelation Coefficient
ACF	Autocorrelation Function
API	Application Programming Interface, a wrapper around (usually) complex functions which can be reused by developers.
BCA	Bias Corrected Accelerated Bootstrap
CEFAS	Centre for Environment Fisheries and Aquaculture Science
CFP	Common Fisheries Policy
CNN	Convolutional Neural Network
CPUE	Catch per Unit Effort
CS	Citizen Science
CSS	Cascading Style Sheet
DCF	Data Collection Framework
DAK	Digitally Accessible Knowledge
EIRE	Republic of Ireland
EU	European Union
FoV	Field of View
FMM	FishMap Môn Survey
GES	Good Environmental Status
GUI	Graphical User Interface
HTTP	Hypertext Transfer Protocol
HTML	Hypertext Markup Language
ICES	International Council for the Exploration of the Sea
IoU	Intersection over Union
IQR	Inter Quartile Range
IRA	Inter-rater Agreement
LK	Local knowledge

MAE	Mean Absolute Error
MBE	Mean Bias Error
MCRS	Minimum Conservation Reference Size, term which superseded Minimum Landing Size (MLS) in the EU.
MHW	Mean High Water
MRF	Marine Recreational Fishing/Fisher
MRIP	Marine Recreational Information Program
MSFD	Marine Strategy Framework Directive
MSP	Marine or Maritime Spatial Planning
MV	Machine Vision
NLP	Natural Language Processing
NOAA	National Oceanic and Atmospheric Administration
OP	Opening Post
OS	Ordnance Survey
PACF	Partial Autocorrelation Function
PAEK	Polynomial Approximation with Exponential Kernel
R-CNN	Regional Convolutional Neural Network
REM	Remote Electronic Monitoring, using digital recording devices to record some activity
RF	Recreational Fishing/Fisher, includes both marine and freshwater activity.
RMSE	Root Mean Square Error
RoI	Region of Interest
RWLPP	Real-world Length per Pixel
SA2012	Sea Angling 2012
SaaS	Software as a Service
SM	Social Media
SNS	Social Networking Site, refers to the giant social networks such as Twitter, Facebook, Instagram etc.

T&C	Terms and Conditions
TDM	Text and Data Mining
TF	Tensorflow
TL	Total Length
UK	United Kingdom
UKHO	UK Hydrographic Office
UGC	User Generated Content
URL	Uniform Resource Locator
USD	United States Dollars
W3C	World Wide Web Consortium
WAM	Wales Activity Mapping Survey
XML	Extensible Markup Language

1 General Introduction

1.1 DEFINITION

Many definitions of marine recreational fishing (MRF) exist (e.g. EIFAC, 2008; ICES, 2013; Pawson et al., 2008). The ICES Working Group on Recreational Fisheries Surveys defined MRF as "*the capture or attempted capture of living aquatic resources mainly for leisure and / or personal consumption, and covers active fishing methods including line, spear, and hand-gathering and passive fishing methods including nets, traps, pots, and set–lines*" (ICES, 2013b). Some definitions exclude subsistence fishing and fishing where the catch is sold or otherwise traded for export, domestic or black markets (EIFAC, 2008; Pawson et al., 2008). The term "recreational fishing" (RF) is synonymous with angling (Pawson et al., 2008) however, angling is defined as fishing with hand lines or fishing rods using baits or artificial lures and other methods exist (ICES, 2013b). Nevertheless, angling has been found to be the dominant method—where data are available—in the European Union (review Hyder et al., 2018) and the Anglosphere countries (e.g. Giri and Hall, 2015; Lovell et al., 2013), although this can be species specific (e.g. Ministry for Primary Industries, 2012). Spearfishing may be comparatively more common where water temperatures are elevated, but angling still tends to dominate throughout Europe according to the current best available knowledge (review Hyder et al., 2018). Passive methods are also used (e.g. seine and fixed nets, traps and set-lines) however there has been little or no focus on surveying such methods, probably because the use of passive methods is comparatively low (e.g. Armstrong et al., 2013a). For the purposes of this thesis, MRF follows the ICES (2013a) definition. The terms angling and fishing will be encountered in this thesis; the terms are intentionally used according to their definition. The abbreviation MRF is also used and for the sake of convenience MRF may be taken to mean marine recreational fishing or marine recreational fisher according to the context. Marine recreational fisheries is not abbreviated. This principle also applies to recreational fishing (RF).

1.2 IMPORTANCE

1.2.1 PARTICIPATION AND ECONOMICS

Many post-industrial countries have economically valuable marine recreational fisheries. Examining 5 of the countries ranked in the top 10 by gross domestic product gives MRF participation estimates of ~ 21 m and direct annual expenditure at ~ 30 bn USD (Armstrong et al., 2013a; Barrow, Brickley, Dumbrell, Johnson, & Fisheries and Oceans Canada, 2012; Henry & Lyle, 2003; Herfaut, Levrel, Drogou, Thébaud, & Véron, 2012; Indecon, 2007; Nautilus Consultants, 2000; A. Radford, Riddington, & Gibson, 2007). Across the EU, the total economic impact was estimated to be ~6 bn euro, which provided support for 100,000 jobs (Hyder et al., 2017). Both Cooke and Cowx (2004) and Arlinghaus et al. (2015) estimated global participation rates to be approximately 11% (saltwater and freshwater). Within the UK there are an estimated 1.08 m participants who engage in sea angling annually, which provides support for 10,400 full time jobs (Armstrong et al., 2013a). UK MRFs were estimated to catch 10.1 m fish annually during 3.8 m fishing days. MRF also has a positive economic impact on income and employment in coastal communities by increasing visitor frequency (A. Brown, 2012; TNS Global, 2014a, 2014b).

The average age of MRFs in the UK was 51 years, with an estimated 98% of participants being male (A. Brown et al., 2013) and this pattern is generally repeated—with some variation—in other developed countries (Barrow et al., 2012; Giri & Hall, 2015; Murdock, Loomis, Ditton, & Hoque, 1996; Ryan et al., 2015). However, there are some indications that participation in angling (saltwater and freshwater) has been decreasing over the last ~3 decades in the UK, Australia and some American States (Barrow et al., 2012; Dann, Alvarado, Palmer, Schroeder, & Stephens, 2008; Giri & Hall, 2015). There is insufficient data to assume this is a general trend across developed countries. Understanding why fishers fish is important if the angling sector is to continue to maximise benefit provision and for management and planning (Arlinghaus, 2006; Arlinghaus et al., 2015; Murdock et al., 1996) and MRF can make significant economic and social contributions in coastal areas where commercial fishing has suffered a decline (Arlinghaus & Cowx, 2002).

1.2.2 Non-economic Benefits

RF benefits participants across multiple ages and economic classes by directly providing exercise and relaxation, and through more esoteric benefits at social and community levels (collectively coined biopsychosocial). Assessing the benefits of participation requires qualitative approaches that can measure the individual benefits from participation and the wider social and community benefits (A. Brown, Stolk, & Dojhari, 2010). Current assessment approaches use survey techniques employed in the health and social sciences (Griffiths, Bryant, Raymond, & Newcombe, 2016). Motivations for MRF are not necessarily related to catching fish (A. Brown et al., 2013; Stolk, 2009) and survey respondents report that spending time with friends and relations, relaxation, physical activity and being in the natural environment are also important (A. Brown et al., 2013; Drew Associates Ltd., 2004; Kenter et al., 2013; Lawrence & Spurgeon, 2007; Mawle & Peirson, 2009). These drivers are also common to anglers in other countries, including Australia (Frijlink & Lyle, 2010; McManus, Storey, & White, 2011) and the USA (Gartner, Love, & Erkkila, 2002).

1.2.2.1 Society

Sea angling is predominantly a social activity with < 20% of MRFs engaging in solo trips in the UK (A. Brown et al., 2013) and fishing was shown to be important for social affiliation in Canadian freshwater fisheries (Gartner et al., 2002). Sea angling promotes social cohesion by bringing together different societal groups with ~33% of MRFs befriending people from different backgrounds, age groups and economic classes (A. Brown et al., 2013; Indecon, 2007). Angling can provide good disability inclusion by providing access to angling opportunities where participation by disabled people can be as high as 20% (A. Brown, Djohari, & Stolk, 2012; A. Brown et al., 2013). Participation rates in angling by people with disabilities can exceed that of other sports (Indecon, 2007). The social and psychological benefits can be significantly greater for physically disabled anglers than those who are able-bodied (Freudenberg & Arlinghaus, 2010). Qualitative evidence also suggests that angling can be used as a tool to engage socially excluded young people, reducing anti-social behaviour and crime (A. Brown et al., 2012; Indecon, 2007).

1.2.2.2 Health

The extended duration of angling activity (> 5 hours average trip duration) means that—despite activity generally being low or moderate—it represents a total energy expenditure that is comparable to mountain biking (Pretty, Peacock, & Hine, 2007) hence participation brings the benefits associated with physical activity (Lawrence & Spurgeon, 2007). Health and fitness is important in tackling obesity and has become an issue across industrialised countries with high levels of obesity. 65% of MRFs rated their physical activity as moderate or high (A. Brown et al., 2013), suggesting that angling could be important in promoting exercise. Angling can build resilience to ill health and improve recovery from both physical and mental illness (McManus *et al.* 2011). These effects are probably mediated through participants feelings of relaxation, reduced stress and increased physical activity (McManus et al., 2011; Ormsby, 2004). In Australia, angling was seen to improve health and wellbeing particularly through stress relief and relaxation, but also through family bonding (McManus et al., 2011).

1.2.2.3 Environmental Benefits

Angling can encourage participants to engage with the natural environment, raising their awareness of marine environmental issues. MRFs have contributed to scientists' understanding of fish and fisheries by participating in many hundreds of volunteer research programs worldwide (e.g. Billfish Foundation, 2018; Environment Agency, 2014; Fairclough et al., 2014; Inland Fisheries Ireland, 2018; Shark Alliance, 2015). MRFs also engage in environmental improvement initiatives, including beach cleans (A. Brown et al., 2013) and campaigns to remove litter (Angling Trust, 2015). MRFs report suspected illegal fishing activity and other events which may negatively impact the environments in which they fish (NRW, Welsh Government Fisheries Dept. pers. comm.).

1.3 Impacts

With echoes of the views held on commercial fishing at the end of the 19th Century[1], MRF has been perceived as having little impact (commentary Lewin et al., 2006). Even in closed

[1] Thomas Huxley, UK Royal Commission on Sea Fisheries 1866, "it is inconceivable that the great sea fisheries, such as those for cod, herring and mackerel, could ever be exhausted".

freshwater systems, comparatively recent research has suggested that RF can be self-regulating (Carpenter, Munozdelrio, Newman, Rasmussen, & Johnson, 1994; Hansen, Beard, & Hewett, 2000; B. D. Smith, 1999). Nonetheless, RF can have a meaningful impact on fisheries and associated ecosystems (Coleman, Figueira, Ueland, & Crowder, 2004; Cooke & Cowx, 2004; Lewin et al., 2006; McPhee, Leadbitter, & Skilleter, 2002). In particular, MRF has been implicated in failure of stock recovery programs (Sherwood & Grabowski, 2016), ecosystem collapse (Altieri et al., 2012), negative population structure effects (Schroeder & Love, 2002; Westera, Lavery, & Hyndes, 2003) and environmental damage (Asoh, Yoshikawa, Kosaki, & Marschall, 2004; Chiappone, Dienes, Swanson, & Miller, 2005). Although all these impacts are important, the dominant concern tends to be the potential deleterious effects MRF may have on species which are also both subject to commercial pressure and important to the commercial sector. Such species are likely to be valued by both commercial and MRFs where stocks are accessible to both groups. Across the EU several stocks have significant MRF induced mortality, with removals comparable to mortality induced by the commercial harvest. This is despite release rates > 70% for some species and in some EU marine recreational fisheries (Armstrong et al., 2013a; Ferter et al., 2013; Weltersbach & Strehlow, 2013). Post-release mortality is fishery dependent with mortality rates ranging between 0-95%, with main explanatory factors of species, bleeding, hooking location, temperature and handling time (A. Bartholomew & Bohnsack, 2005; Lewin et al., 2018; Weltersbach & Strehlow, 2013). Several articles are in agreement that recreational biomass removals are high (e.g. Armstrong et al., 2013a; Herfaut et al., 2010). The review of Radford et al. (2018) put recreational removals at 26% for European sea bass in ICES divisions VIIa-h and IVa-c. Atlantic cod removals were also 26% in divisions IIIb and IIIc and Atlantic pollock removals in divisions VIIa-h were estimated to be 43%.

To date, estimates of MRF harvest have not been widely used in stock assessments and management of marine stocks under EU control, in spite of significant recreational harvest (Z. Radford et al., 2018). However, as reporting requirements under the revised data collection framework (DCF) are met, the situation could be expected to improve. Examples of recreational harvest being included in EU stock assessments include Atlantic salmon, European sea bass, Atlantic cod, salmon and sea trout (ICES, 2014b, 2017c). Including recreational catch in species with comparatively high recreational removals will support the management of important stocks.

1.4 Management and Policy

1.4.1 Licensing

Within the UK, areas used by MRFs are de facto “commons” (definition, Hardin, 1968). No license is required to participate, access to waters and the intertidal zone is almost completely unrestricted, and there are no harvest controls in place for the majority of species. The picture across the rest of the EU varies between member states, with some countries requiring no license (e.g. EIRE), or other countries requiring licenses based on activity e.g. for coastal trolling and net fishing in Sweden (Hyder et al., 2018). A summary of EU recreational licensing is given in Hyder et al. (2017). The lack of any (or incomplete) registration or licensing schemes makes the design of surveys more difficult as there is no precompiled sampling frame from which to draw random samples (Ashford, Jones, & Fegley, 2009; National Research Council, 2006; Pollock, Jones, & Brown, 1994). Licensing or registration schemes also provide lists of fishers from which volunteer diarists can be drawn, and importantly measures can be taken to control any bias arising from refusals from selected diarists.

1.4.2 Reporting and Policy

In the European Union the DCF (European Commission, 2016a, 2017b) requires member states to collect data on recreational harvest and releases of Atlantic cod (*Gadus morhua*), European sea bass, Atlantic pollock (*Pollachius pollachius*), Atlantic salmon (*Salmo salar*), European eel (*Anguilla anguilla*), sea trout (*Salmo trutta*), and some elasmobranchs and tunas. Not all member states have to collect data on these specie sand member states who have provided evidence that catches are below a threshold receive a reporting derogation. Collection of MRF data in the EU is also required under Control Regulation 1224/2009 (European Commission, 2009) that requires MRF is “conducted in a manner compatible with the Common Fisheries Policy” and that recreational catches of fish subject to EU recovery plans are monitored where catches are made from boats (shore fishing is currently excluded). Regulation 1224/2009 also mandates the European Council to apply management measures to marine recreational fisheries if research by the Scientific, Technical and Economic Committee for Fisheries indicates that recreational harvest is impacting the stock.

The EU can implement regulations for specific stocks which have legal jurisdiction and are legally binding across all member states however, countries will have different national and local regulations which are in place to ensure the country meets EU fisheries directives or laws or meet local needs. Devolved regulatory agencies with a legal responsibilities for marine management and fisheries can apply technical measures to limit MRF activity.

There are several other EU directives—and hence a plethora of national level regulations—which require assessment and potential ongoing monitoring of MRF activity, but the interaction of relevant instruments is opaque (Boyes & Elliott, 2014). In 2014, Directive 2014/89/EU (European Commission, 2014) was ratified by the EU, thereby creating a framework for marine (alt. maritime) spatial planning (MSP). MSP seeks to achieve the equitable allocation of marine and coastal resources where stakeholder activities are potentially in conflict. The aim of using an MSP framework is to ensure that benefit maximisation occurs now and in the future and MSP is considered a vital component of ecosystem-based management (Douvere, 2008; Environmental Law Institute, 2009). The concept has been adopted internationally, and MSP policy is not limited to EU member states (e.g. The White House, 2010; Vince, 2014). MSP policy recognises that pressures on marine areas and resources are likely to increase with increasing demand and MSP seeks to; (i) manage competing pressure within an ecosystems-based management paradigm; (ii) maximise the benefits realised by compatible activities and (iii) integrate with terrestrial planning. It is important that MRFs are not disenfranchised from any processes because of a lack of information on their activity.

Member states must also consider all aspects of MRF activity which have a bearing on the Marine Strategy Framework Directive (MSFD; European Commission, 2008). The MSFD is the legislative instrument which requires member states to achieve Good Environmental Status (GES) by 2020 according to the criteria in Commission Decision 2017/848 (European Commission, 2017a). Although the relationship between MSP and MSFD is unclear (Boyes & Elliott, 2014; Brennan, Fitzsimmons, Gray, & Raggatt, 2014). Descriptor 3 of Commission Decision 2017/848 is the most pertinent to the assessment of marine recreational fisheries, it states that "*populations of commercially-exploited* fi*sh […] are within safe biological limits [and the stock is healthy]*" and that "*the fishing mortality rate […] is below levels which can produce the maximum sustainable yield*". This implies that mortality from recreational harvest both known and insignificant. The Decision also covers anthropogenic effects on ecosystems, benthic habitats and marine litter which may be relevant to MRF at local scales e.g. where high concentrations of boats set anchor on reefs to target fish aggregations during spawning.

1.4.3 Management Measures

Actions may be necessary to protect localised marine ecosystems (CFP, MSFD, MSP), reduce conflict between marine users (MSP), reduce impacts on other species (CFP, MSFD), reduce harvest (CFP, MSFD) or to address non-marine issues such as trespass or noise disturbance (MSFD, MSP, Other). Technical measures to limit annual harvest includes changing (or introducing) limits on the total length of fish which can be retained (i.e. slot sizes or minimum conservation reference sizes, MCRS), introducing bag limits, "gear" restrictions (e.g. a ban on live-baiting with sandeel), and temporal / spatial closures of marine areas. Multiple technical measures have been applied to the recreational European sea bass fishery in the North East Atlantic following estimates of harvest above maximum sustainable yield in 2013 (ICES, 2013a). Harvest controls were put in place, including an increase in the MCRS to 42 cm in September 2015 (commercial and recreational) in ICES divisions IVb-c, VIa-b, VIIa-d, VIIh, VIIj (European Commission, 2015a). Bag limits have been in place in ICES divisions IVb-c, VIIa, VIId-h, VIIj-k for recreational anglers since March 2015 (European Commission, 2015b), with zero bag limits imposed for the periods January 2016 – June 2016, January 2017 – June 2017 and January 2018 – September 2018.

1.5 Novel Approaches

Data collection under statistically sound directed survey programs will remain essential for the provision of scientifically credible indicators of the state of recreationally targeted stock and related activity. However increasing restrictions on public budgets, coupled with increasing legislative demands to assess MRF impacts has highlighted the need to use new innovative methods to collect fisheries data that are cost-effective and can meet legislative requirements. Existing datasets on MRF activity tend to be fragmentary because cost and logistics prevents their frequent and regular execution. This can lead to fragmentary datasets across space and time which may reduce confidence in interpolated estimators of catch and effort (Griffiths et al., 2014). The burden of reporting on marine recreational fisheries is expected to increase in the future, and where activity does not justify a reporting derogation, regular assessments will need to be carried out.

1.5.1 "Traditional" Survey Sampling in Brief

Traditional survey sampling methods in RF have been well covered in the literature (e.g. Brick et al., 2012; McGlennon and Kinloch, 1997; National Reseach Council, 2006; Pollock et al., 1994; Steffe et al., 2008; van Voorhees et al., 2002). In brief, population estimator of total harvest typically use multiphase instruments. Population estimates of total effort require a comparatively large number of samples across large spatial scales and these are usually gathered by a low cost method (to reduce variance) e.g. random direct dialling, mail survey or by incorporation into an existing national survey program. A second phase is used to obtain catch per unit effort (CPUE) estimators, according to some predetermined stratifications of sampling units which are matched to stratifications in the population effort survey assessment phase. The CPUE phase usually has an on-site component however, diary like approaches are being increasingly used (ICES, 2017d, 2014b, and other working group reports). Diary volunteers are still necessarily self-selecting and good practice would—as a minimum—provide evidence that diary volunteers were not significantly different from non-volunteers. Complementary surveys are usually undertaken to correct for biases (e.g. frame errors and non-response; Pollock et al., 1994), for cross validation (e.g. Holdsworth et al., 2018), and increase statistical confidence in population estimators (Pollock et al., 1994). In reality, each phase usually has more than a single survey instrument and additional corrections (e.g. for length of stay bias) are also applied according to prior knowledge of the direction and magnitude of expected biases. Further complexity is added because different surveys are required according to platform (i.e. charter boat, private boat and shore angling) however, further details are beyond the scope of this thesis.

Complementary survey approaches can help reduce costs and provide data which may be otherwise difficult to obtain (e.g. during the night). Aerial surveys have been used where effort is spread over a wide geographical area with many access points (Smallwood, Pollock, Wise, Hall, & Gaughan, 2012; Vølstad, Pollock, & Richkus, 2006). Remote electronic monitoring (REM) using various approaches have also been used (Hartill, Payne, Rush, & Bian, 2016; Parnell et al., 2010; Steffe et al., 2008; van Poorten, Carruthers, Ward, & Varkey, 2015), but such methods are dependent on limited access points (e.g. harbour entrances and slipways) or to specific venues (e.g. breakwaters). Even though random stratified sampling can be integrated into such methods, care is required as to which sampling units would be excluded from the notional frame from which REM observations are being sampled. Efforts would then need to be made to sample from the excluded frame, assuming some estimator(s) of a wider population is to be calculated.

Marine recreational surveys have tended to be infrequent or non-existent (with exceptions, e.g. MRIP survey in the USA; National Oceanic and Atmospheric Administration, 2017).

Hence MRFs and the stocks they prosecute generally lack historical baselines. This is problematic as it is the patterns of temporal and spatial changes in stock structure which give insights into the response of stocks to environmental pressures and anthropogenic pressures, and also allows the assessment of changes in fisher behaviour (Hilborn & Walters, 1992; Pauly, 1995). Unconventional data has been used to reconstruct historical time series of catch data (e.g. Schiller et al., 2015, 2013; Smith and Zeller, 2015; Zeller et al., 2011, 2007) and sources have included images (e.g. McClenachan, 2009), magazines and newspapers (e.g. McClenachan, 2009; Richardson et al., 2006), and interviews with fishers, book keeping, and logbooks / diaries (e.g. Belhabib et al., 2016).

1.5.2 SOCIAL MEDIA AND OTHER FISHER KNOWLEDGE SOURCES

Social media (SM) use is predicted to rise to 3.02 bn people world-wide by 2021 (Statista, 2018) and every day 0.7 bn comments and 300 m photos are posted to Facebook (Forbes, 2018). This represents an unprecedented amount of data and excludes other SM sites. Some proportion of this content will be open text and images containing records of MRF activity. Historical content is also available from blogs, newsgroups, bulletin boards and discussion forums which predate the social networking giants.

Social media has been harvested for trend data on our behaviours since the first item was sold through a website. But TDM has been used to understand real-world events, including the tracking of flu outbreaks (Dugas et al., 2013), the progress of forest fires (De Longueville, Smith, & Luraschi, 2009) and the impacts of earthquakes (D. Yates & Paquette, 2010). Wilde and Pope (2013) and Martin et al. (2012) used Google Trends (then Google Insights for Search) to attempt to gain insights into RF activity. Google Trends returns a normalised number derived from the number of times a specified search term has been submitted to Google's search engine. Google Trends allows results to be aggregated at different spatial and temporal levels and by general subject categories (e.g. hobbies and leisure). Wilde and Pope (2013) identified changes in angler numbers by country, but only provided a philosophically discursive case for the reliability of the observed trends. Similarly, Martin et al. (2012) looked at search terms associated with angler recruitment to identify the terms most used by anglers associated with retention and recruitment programmes.

Google Trends is one of many analytics tools which report metadata or amalgamated data from users' internet behaviour (e.g. network relationships and sentiment). Content directly generated by users is more commonly used. Content is typically open text, images, audio or video. Both Martin et al. (2014) and Shiffman et al. (2017) extracted information from online discussion forums. Martin et al. (2014) used the count of posts by water body name to show that effort recorded by a bus-route survey correlated with post counts. Whereas Shiffman et al. (2017) manually extracted data from open text and images to gain insights into the behaviour of shark anglers, revealing some illegal practice. YouTube videos have also been used to extract meaningful information. Belhabib et al. (2016) estimated the length of captured fish length and indicators of effort to produce population estimators of total catch and economic value from RF tourism. Banha et al. (2017) successfully demonstrated that social media data—in combination with other sources—could be used to map the spread of the invasive catfish *Ictalurus punctatus*. In Giovos et al. (2018), YouTube movies were used to create a better understanding of the marine recreational fishery in the Mediterranean, including gear use and captured species. Despite increasing support for the automated scraping and processing of open text data, it is surprising that only Google Trends, search term counts or the manual extraction and collation of data from viewed content have been used. This severely limits the volumes of data that can be reasonably processed.

Online digital sources of fisher knowledge are easy to access however, other sources have been used to derive descriptors of recreational fisheries. The knowledge of commercial fishers has been recognised as a potentially vital component in the assessment and management of commercial fisheries (Hind, 2014, 2015; Johannes, Freeman, & Hamilton, 2000; Stephenson et al., 2016) and a comparatively strong body of research currently exists (e.g. Anuchiracheeva et al., 2003; Canese and Bava, 2015; Close and Hall, 2006; Freitas et al., 2009; Hall and Close, 2007; Hamilton et al., 2012; Kafas et al., 2017; Macdonald et al., 2014; Shepperson et al., 2014). However, methods which exploit heterogeneous sources of fisher knowledge to produce descriptors of marine recreational fishers are less common. All surveys of recreational catch are fisheries dependent—researchers do not pick up a set of standardised gear and a sampling protocol sheet then head to the sea to perform fisheries independent sampling. Excluding the use of social media sources of fisher knowledge in research, four articles used sources of fisher knowledge. Two extracted trophy fish records from printed media to demonstrate a significant negative trend in the size of (some) trophy fish (Bellquist & Semmens, 2016; Elizabeth A Richardson et al., 2006) and two used club records to calculate time series of CPUE and evaluated the effect of harvest controls (Bennett, Attwood, & Mantel, 1994; Gartside, Harrison, & Ryan, 1999). In addition, Belhabib et al. (2016) used logbooks to produce estimates of catch from tourist based MRF trips for some countries.

1.5.3 CITIZEN VOLUNTEERS AND APPS

Recreational fishing surveys are increasingly using diaries to replace on-site methods in the calculation of catch composition, catch rates and release rates (Georgeson et al., 2015; Holdsworth et al., 2018; ICES, 2017d). Volunteers are ideally recruited to take part in the survey by some randomisation process typically from licence lists, or from respondents contacted during a separate survey phase. Diaries are maintained by participants according to a format predefined by the survey design to ensure the necessary data is recorded correctly. Trip data is not necessarily recorded during the fishing session which can give rise to recall biases, which is probably the largest single biasing effect among many others (review Bolger et al., 2003).

Citizen science (CS) projects also rely on volunteers and the distinction between diarists as above, and a citizen scientist is indistinct. The definition of "volunteers with no formal training in science collecting, categorizing, transcribing, or analysing scientific data" (Bonney et al., 2014; Silvertown, 2009) does not provide clarification. However, differentiating factors include recruitment methods, information transfer and project longevity. In addition, "citizen science" conjures difficult to define impressions of larger scales (participant count, spatial, temporal) than "volunteer". The largest difference is in the approach to post-collection data processing, with survey diarists' data being part of a predefined survey methodology. Examples of CS projects involving MRFs including tagging (Billfish Foundation, 2018; Shark Alliance, 2015), collection of biological samples (Fairclough et al., 2014; Williams, Holmes, & Pepperell, 2015) and underwater monitoring of fish assemblages (Florisson, Tweedley, Walker, & Chaplin, 2018). The review of Hyder et al. (2015) summarises recent marine related CS based projects.

In 2017 global ownership of smartphones was estimated to be 59% (Pew Research Center, 2018) and adoption is expected to increase in emerging countries (Poushter, 2016). Smartphones can greatly reduce the inconvenience of recording data because they include multiple sensors including GPS, accelerometers, gyroscopes, microphones, cameras and a clock. Water resistance is also increasingly common. These properties make smartphones ideal devices for recording data and several angling smartphone applications (henceforth, simply *angler apps*) have been created. Angler apps allow users to engage in within-app social networking and information sharing. Additional functions include species identification, access to regulation information (e.g. landing sizes) and diary-like catch recording (review Venturelli et al., 2017). Angler apps have been used to map the temporal and spatial distribution of effort

(Papenfuss, Phelps, Fulton, & Venturelli, 2015), making the synergy these angler apps could have with CS and directed surveys clear. It could be argued that angler apps are *de facto* CS projects (but may be described as crowd sourced). Nonetheless angler apps have the potential to collect vast amounts of data from recreational fishers anywhere in the world (assuming offline operation is supported) provided that owners are both prepared to share data and are legally able to share data with third parties.

1.5.4 Remote electronic monitoring

Remote electronic monitoring has been shown to be cost effective and reliable (Blight & Smallwood, 2015; Wise & Fletcher, 2013) and can capture night-time fishing activity where this occurs in a lighted area, or thermal imaging cameras can be used. Cameras can be used opportunistically, e.g. where there are pre-existing web and security cameras (Hartill et al., 2016), or they can be installed temporarily at choke points, specific venues or at key vantage points (Keller, Steffe, Lowry, Murphy, & Suthers, 2016; Lancaster, Dearden, Haggarty, Volpe, & Ban, 2017; Parnell et al., 2010). Remote electronic monitoring (REM) is primarily used to determine effort, effort indices or to monitor activity in protected areas (Lancaster et al., 2017). The nature of the deployment will dictate the limitations, e.g. REM of access points (e.g. harbour entrances and slipways) cannot be used to determine time spent at an exact fishing location. A review of the potential applications of REM is given in the report by Steffe et al. (2017).

1.6 This work

Regular assessment of MRF is a legislative requirement across EU member states, and in other countries (National Oceanic and Atmospheric Administration, 2007). To meet these requirements, impacts on commercially important stocks and ecosystems must be continually assessed where those impacts are above a designated threshold. If there is sufficient evidence that impacts are below the threshold then a derogation may be obtained (European Commission, 2008, 2016a, 2017b, 2017a). As EU citizens, MRFs should also have representation in any MSP processes, and their activity should be assessed within any planning and ongoing management processes (European Commission, 2014). However, the assessment of marine recreational fisheries is costly, hence complementary methods are being sought to ensure that there is an evidence base sufficient to meet all marine legislative requirements and deliver good stewardship. Technology will also bring opportunities to improve the efficiency of "traditional" survey assessment methods and will find wider applications in the monitoring of commercial fisheries and in other ecological disciplines.

Recent novel approaches frequently require the collection and processing of large volumes of data which is not well tabulated, or does not otherwise encode the required data in a prescribed format from which information is readily extracted (e.g. open text and images). Historical data represent an opportunity to extract meaningful content on the past and present activity of marine recreational fishers, which may present the only prospect to create baselines for RF activity (Hilborn and Walters, 1992, Pauly, 1995. However, there are difficulties and the sheer volume of data may require the use of emerging techniques in machine learning and natural language processing to efficiently extract the required information. These difficulties are particularly relevant where recreational fisheries dependent information is being reported in the following scenarios; (i) social media; (ii) use of smartphone applications by fishers for personal use and (iii) in volunteer lead science. Each of these scenarios can embed activity data in multiple media types (e.g. open text and images). Hence this thesis introduces methodological approaches to data gathering and data processing of in data forms commonly encountered on social media and also contained in other recreational fisheries dependent data sources.

To date, there has been little or no use of automated processes in the extraction and processing of data in marine recreational fisheries research. This is despite the increasing interest in these alternative sources to describe marine recreational fisheries. As the volumes of available data increases, manual methods become increasingly impractical (e.g. Martin et al., 2014, Belhabib et al., 2016, Shiffman et al., 2017), hence fisheries researchers will need to adopt modern computational techniques in their data processing pipelines. The overarching theme of this thesis is to show the application of methods which demonstrate how a basic set of programming skills—now common held by many researchers in the ecological sciences—can be used to exploit recently published and accessible application programming interfaces to perform comparatively complex data processing tasks which can be used to automate the extraction of meaningful information from different source and types of fisher dependent data. An emphasis is placed on validating the approaches against data derived from accepted methodologies to contribute to the growing body of evidence that non-traditional approaches are valid. This is also the justification for publishing the chapters as articles to scientific journals, ensuring the research is disseminated effectively.

Chapter 3 introduces a comparatively unsophisticated and custom approach to the extraction of data from a large volume of open text. The thesis culminates in Chapter 6 by demonstrating how newly developed advances in machine learning can be applied to that most fundamental of fisheries assessment tasks—the estimation of fish length. The convolutional neural network approach used in Chapter 6 to automate fish length estimates from images has also been deployed in optical character recognition, acoustic analysis and natural language processing and is not only applicable to the application presented herein. It is apparent that each of these uses has the potential to be employed in many different scenarios applicable to the assessment of recreational fisheries and other ecological research. Individual chapter introductions and discussions provide additional detail and context.

In using any historical sources of fisher knowledge, there are ethical considerations surrounding the use of the fishers' data. This holds true even when data were not solicited from an individual by the researcher and is particularly pertinent to SNS's where institutional ethical policy may be unclear. Where user data is retained and that data would allow the individual to be identified, then the research probably falls under the definition of human subjects research (World Medical Association, 2001). Human subjects research necessitates obtaining informed consent from the individuals who can be identified. There are other ethical aspects to consider, including the balance of research quality and the possible maleficience and beneficence to individuals who may be affected in conducting the research or by the outcomes of the research. Concerns can also arise regarding copyright, fair-dealing and lawful access, particularly with ethics committees or peers who have had little exposure to the text and data mining (TDM) of SM. The application of data mining and machine learning provides new opportunities to extract historical data from sources of fishers' knowledge which may not have been previously practical to exploit, hence it is important to avail potential researchers with an accessible treatment of the relevant legal and ethical issues which they may face.

1.6.1 Objectives and Questions

This thesis seeks to introduce novel methodologies in two areas of marine data collection, fisher knowledge and images. Several chapters use the European sea bass (for brevity, simply sea bass) as a test subject and Appendix K gives an overview of the species. Chapters 2, 3 and 4 are predominantly concerned with the extraction of data from fisher knowledge sources with a particular emphasis on open text data published to SM. Prior to engaging in any passive SM based research it is important for researchers and students to be aware of the ethical issues which surround the use of user contributed information to SM.

Chapters 5 and 6 look specifically at the length estimation of fish from images using a foreground fiducial marker. Notably, Chapter 6 uses machine vision (MV) to detect an object in an image, which has applications to many areas in the assessment of recreational and commercial fisheries. Images and videos (images) are recorded by RFs or may be used in the assessment of RF activity as follows; (i) Published by users to social media and media sharing platforms; (ii) Captured on smartphones during the use of angler apps; (iii) Captured using angler apps designed for the diary-like phases of directed surveys; (iv) Used in REM to record CPUE and stock population structure on charter boats and commercial boats; v) Used in REM in the assessment of effort (e.g. derive indices of private boat activity by recording departures from harbours).

Chapter 2 gives a review of the issues surrounding the gathering, retention and processing of social media for the purposes of fisheries research however, it is equally applicable to other ecological research fields.

Chapter 3 asks the question; can the partially automated TDM of social media data (specifically online discussion forums) provide meaningful quantitative data on an important subpopulation of MRFs? The shore-based recreational European sea bass (*Dicentrarchus labrax*) fishery of Wales is used as an example fishery and temporal patterns of activity are validated against the Sea Angling 2012 survey data (CEFAS, 2012).

Chapter 4 asks; can different sources of fisher knowledge be combined to produce high resolution maps of the distribution of MRF effort? The results are validated against ground truth data from two other surveys.

Chapter 5 asks; can a mechanistic methodology be used to reduce error and bias in the estimation of fish total length when using single camera photogrammetry with foreground, background and laser fiducial markers?

Chapter 6 asks; can MV be used to produce estimates of total length of the European sea bass from images and how accurate and precise are those estimates to changes of image scale, rotation and flipping? This chapter shows how multiple elements can be combined within an image processing pipeline to produce automatic estimates of total length from an image.

Chapter 2

The Ethics of Using Social Media in Fisheries Research

Chapter 2 was published in the journal *Reviews in Fisheries Science and Aquaculture*.

Monkman, G. G., Kaiser, M., & Hyder, K. (2018). The ethics of using social media in fisheries research. *Reviews in Fisheries Science & Aquaculture*, 26(2), 235–242. https://doi.org/10.1080/23308249.2017.1389854

KH and MK critically reviewed and revised the article. All other work was that of GGM.

No highlights were required by the journal.

2 The Ethics of Using Social Media in Fisheries Research

2.1 Abstract

The use of social media data is becoming increasingly widespread in ecological research and this trend is expected to continue as social media use increases globally. Fishers share details of their activity online and scientists have mined this content to help understand fisher activity, yet little information exists on the ethics of exploiting social media for fisheries research. In this paper, the ethics of using social media published data in fisheries research is discussed in

the context of affected stakeholders and potential causes of maleficience. The legal position with respect to copyright and fair use is summarised in relation to the use of fisher data published on the internet for research. It is argued that research per se does not sensu stricto involve human subjects where no new content is solicited from participants. Text and data mining of social media for research purposes generally receives special dispensation in law to allow scientific endeavour to be conducted without fear of prosecution. Nevertheless, researchers have a professional duty to weigh research benefits against the risk of causing harm to involved agents, including website owners. Ultimately researchers should continually reassess the ethics of their social media research as guidance from ethical review boards currently may be limited and all internet content scraping activity should be conducted responsibly such that personal data is not compromised.

2.2 INTRODUCTION

Currently only a small proportion of the world's marine stocks are sufficiently data rich to enable formal stock assessments to be performed hence most marine fisheries are data poor (Costello et al., 2012; Ricard, Minto, Jensen, & Baum, 2012). This data gap has led to increasing interest in how historical and current fisher knowledge can contribute to the data necessary to deliver effective, equitable and transparent fisheries management policy in otherwise data poor fisheries (Hind, 2015; Stephenson et al., 2016).

Marine recreational fisheries in particular can lack current and historical data even in developed countries (Hyder et al., 2018; ICES, 2017d). Many countries have had little or no catch recording, no registration or licensing schemes for some marine recreational fishing sectors, and no regular national directed surveys. Recreational fishing can have a meaningful impact on fisheries and associated ecosystems (Cooke & Cowx, 2004; Lewin et al., 2006; McPhee et al., 2002). Therefore researchers have sought to exploit novel data sources, including club records (Bennett et al., 1994; Gartside et al., 1999), recreational fisher magazines (Barbini, Lucifora, Figueroa, & Gillanders, 2015; Elizabeth A Richardson et al., 2006), social media (Belhabib et al., 2016; D.R. Martin et al., 2012; Dustin R Martin et al., 2014; Shiffman et al., 2017) and smartphone applications (review Venturelli, Hyder, & Skov, 2017) to better describe data limited fisheries. There is now an increasing adoption of social media as a data resource for ecological research (e.g. Barve, 2014; Daume, Albert, & von Gadow, 2014; Galaz et al., 2010; Joly et al., 2014), following successes in public health monitoring (review Eysenbach, 2011) and crises management (review Alexander 2014).

The ethics associated with conducting human surveys using questionnaires (even when sensitive data is processed) are well understood by ethical review boards (ERB) due to their long-established use in the medical sciences. Typically in fisheries research the use of questionnaire-based surveys requires no recording of personal information and hence readily conforms to recognised ethical standards (e.g. Hall and Close, 2007; Yates and Schoeman, 2013; Léopold et al., 2014; Kafas et al., 2017). In contrast, fisheries research which uses publicly-available social media data will rarely require approval through formal ethical review. This situation places the onus on the researcher and their peers to make their own ethical assessment of any proposed research.

Conversely, it is possible that research proposals based on social media content may be stymied by institutional or peer objections because of confusion over what constitutes human subjects research, the legal position over intellectual property and copyright of published content, and concerns over privacy and data protection. ERBs may require researchers to comply with unnecessarily onerous criteria in conducting social media research which renders such research impractical, an issue raised by Walther (2002) following guidance published by the American Association for the Advancement of Science (Frankel & Siang, 1999).

Ethical aspects of using fisher knowledge have been considered previously (Maurstad, 2002; Silver & Campbell, 2005), but discourses on the ethical use of fisher knowledge published on social media are non-existent. This paper offers insights into the ethics and legalities of mining social media content in fisheries research. Here, fisher published knowledge refers to content which individuals or groups publish online without direct solicitation of that content from a scientist.

2.3 Social Media and Ethics

When conducting research, scientists must meet the ethical guidelines defined by their (i) Funding or commissioning bodies, (ii) Institutional data management department (e.g. IT services) and (iii) Institutional ethical review boards. Research involving human subjects or experimentation on an animal under legal protection for research purposes are controlled under national statute law. The breach of such statutes (e.g. Office for Human Research Protections, 2009; UK Parliament, 1986) can result in criminal prosecution of researchers and their supporting institutions.

When mining social media content, it is clear that no animal is subjected to direct experimental intervention. It may be less clear if social media-based research *sensu stricto* involves a human subject as the discipline lacks the history of the biomedical sciences where bioethics is deeply embedded within research institutions and is an active area of research in its own right as evidenced by the number of journals covering ethics (ScimagoJR, 2017).

Human subjects research must observe four tenets as set out by the Declaration of Helsinki (World Medical Association, 2001) and by Beauchamp and Childress (2001); (i) Beneficence, which requires research to “do the most good”. (ii) Nonmaleficience (“do no harm”), (iii) Respect for persons (participants exercise free will in their continued participation through informed consent) and (iv) Justice (equitability in participant selection and fair distribution of risk and benefit arising from the research). In the context of text and data mining of social media, tenet (iii) would necessitate obtaining informed consent from content authors should the research be classified as involving human subjects. In particular, obtaining consent would be problematic when using automated text and data mining techniques to process large volumes of user generated content (Norval & Henderson, 2017). Informed consent can be difficult to obtain because users can be inactive, use pseudonyms, and personal details are typically protected by the hosting website. In addition, any anonymisation of downloaded data would render it impossible to honour users' requests to withdraw their data from the research. Thus the very nature of social media makes it difficult for a researcher to comply with tenet (iii).

When considering if research involves human subjects, the World Health Organisation's (WHO) definition of human subjects research is of key importance. That definition is "any … systematic collection or analysis of data … to generate knowledge, in which humans are i) exposed to manipulation, intervention, observation, or other interaction with investigators directly or through alteration of their environment, or ii) become individually identifiable through investigator's collection, preparation, or use of biological material or medical or other records" (World Health Organisation, 2015). It is apparent that mining fishers' social media data for research purposes unambiguously falls outside the WHO's human subjects definition when there is no solicitation of a response by the researcher (passive research) and content is anonymised.

The review of Walther (2002) underlines the argument that the use of public internet content in quantitative research does not involve human subjects, provided there is no recording of personally identifiable information and no public reproduction of the original user generated content. Since Walther's (2002) review, the efficiency of web page indexing services (e.g. Google) and advances in big data analysis have increased the ease with which individuals can be identified. These technological advances mean that the researcher must consider carefully what data to publish. For example, by searching online using an excerpt of user generated content it might be possible to identify the original author. In addition, in publishing raw data it has been shown that just a few information points (indirect identifiers) can be used to de-anonymise a user (Narayanan & Shmatikov, 2009). Examples of indirect identifiers (as opposed to direct identifiers) are workplace, occupation and current town of residence. With respect to fisher data, there is the potential to identify a user using a web search from a single published data point, such as the specific weight of a captured fish.

Although it may be possible to identify a content author, the majority of social media based fisheries research would not have any requirement to process or publish personally identifiable information (PII). Incidental collection of personal data may be unavoidable when using automated content scrapers, but this data poses no ethical dilemma provided that; (i) The content was public, (ii) The PII is neither processed nor published and (iii) The original content is securely deleted after processing. The retention and processing of raw data that contains PII is likely to contravene national data protection laws.

Some fisheries research may involve the human as the primary subject, particularly in small-scale and emerging fisheries, or where there is interest in subpopulations of fishers (e.g. Freudenberg & Arlinghaus, 2010). Research using social media to solicit a response from an individual (e.g. questioning respondents in a chat room) could qualify as human subjects research. Researchers must take particular care when collecting and processing special PII which is defined by the European Union as records of race, ethnicity, political opinions, beliefs, trade unions membership, sexual behaviour and health where these data can be associated with an identifiable individual. The European Union has multiple legal instruments which require member states to legislate to prevent the retention and publication of PII and records of criminal activity (e.g. convictions for illegal fishing activity) and minors (European Commission, 1995, 2000, 2010, 2016d). Similar protections exist in Australia (Australian Parliament, 1988), New Zealand (New Zealand Parliament, 1993) and Canada (Canadian Parliament, 1985b), although the situation in the USA is more fragmented with legislation enacted at the state level. Legally it may be possible to process special PII for scientific research in some jurisdictions. Article 89 paragraph 2 of EU regulation 2016/679 (European Commission, 2016d) states "where personal data are processed for scientific … research or statistical purposes, Union or Member State law may provide for derogations [from data protection rights where] such rights are likely to render impossible or seriously impair …the achievement of the specific purposes". Clearly anonymization and data security must be carefully considered in consultation with institutional data management personnel and ethical review board(s) as institutional policy will *de facto* supersede other considerations.

Irrespective of whether research involves human subjects or may contravene data protection law, scientists have a professional duty to ensure their research is non-maleficent (i.e. causes no net harm). As the mining of social media content has no direct physical intervention, assessment of non-maleficience must consider the potential influence exerted by research outcomes on all stakeholders (e.g. local communities, managers and policy makers). Published data could also influence fisher behaviour, for example changing patterns of effort and capture methods which could have localised impacts on targeted fish populations and the environment (e.g. by increasing boat traffic in a marine protected area or publicising more efficient capture methods).

It is probable that institutional guidance on the associated risk of indirect impacts on marine fisheries, and human stakeholders may not exist. Where guidance does exist, it may not provide explicit instruction for social media research scenarios and no national or international ethical frameworks exist for the ecological sciences (Crozier & Schulte-Hostedde, 2015; Markham & Buchanan, 2012; Minteer & Collins, 2005). The researcher must therefore inductively and continually balance the benefit of their work, against the risk of maleficience to stakeholders who are knowingly or unknowingly involved, and the ecosystems to which those stakeholders are linked.

In addition to the direct authors of mined social media other entities also require consideration of ethical issues. These entities are listed below and discussed in the following sections:

i. Social media platform owners and other parties involved in platform hosting services
ii. Researchers
iii. Human stakeholders for which the ecological resources have social or economic value
iv. Species, ecosystems and environments

2.3.1 PLATFORM OWNERS AND HOSTING SERVICES

Source web sites usually have terms and conditions (T&C), and acceptable use policies which allow for the use and reuse of user generated content. Despite this, reputational damage to a web site can arise when users perceive the space in which they publish as private but find that such privacy is an illusion when research using their content is published. Even though users may not hold the site directly responsible, users may no longer be willing to freely contribute content knowing it can be used outside of their online community and for scientific purposes. These considerations present a dilemma for researchers who can choose to retain the anonymity of content sources to reduce the risk of reputational damage to websites, but only at the cost of methodological transparency.

Legal exceptions to copyright law are present in many jurisdictions (termed fair use or fair dealing) which allow for the use of copyrighted material in non-commercial research provided the content is lawfully accessed. European Directive 2001/29/EC (European Commission, 2001) instructs member states to make legal provision for copyright exceptions for research. Further examples of fair use legislation for research purposes are the United States (United

States Code, 2011), Australia (Australian Parliament, 1968), Canada (Canadian Parliament, 1985a) and New Zealand (New Zealand Parliament, 1994). More generally, 174 countries are signatories to the Berne Convention (World Intellectual Property Organization, 1979), an international agreement on copyright law between countries. The convention states that exceptions to copyright protection are allowable where there is no conflict with the normal use of copyrighted material and where the interests of the right holder(s) are not prejudiced.

Text and data mining (TDM) of online material is a special case because it involves the verbatim copying and storing (usually transitory only) of complete texts or audio-visual media. There is virtually no specific legal provision for TDM exceptions (including in the Berne Convention) however, United States case law has multiple examples of the courts upholding fair use for TDM (Cox, 2015). There are indications that specific legal exemptions for TDM will be codified. For example, the United Kingdom already provides explicit exemptions from copyright for TDM in the Copyright, Designs and Patent Act 1988 S. 29A (HMSO, 2016) and the European Union is currently developing a new directive covering TDM (European Commission, 2016c). In both cases data must be lawfully accessible for the exception to apply for non-commercial research.

As there is no definition of lawful access with respect to TDM in Directive 2016/0280(COD) it is reasonable to assume that all publicly accessible web data is lawfully accessible and that this extends to data held behind a subscription paywall. The United Kingdom's 2016 revision to the Copyright, Designs and Patent Act 1988 goes further and there is explicit legal protection preventing a website's T&Cs stopping fair use. The legislation further adds that when web data is lawfully accessed, publishers should not stop a researcher from mining data (HMSO, 2016).

Content mining activity can negatively impact site performance, slowing response times and in extreme circumstances, causing the site to become unavailable to users. Smaller scale social media sites are typically hosted by third party service providers who supply the software and hardware infrastructures necessary to operate the service. These infrastructures frequently share resources across many services, hence automated scraping of published content can impact many parties unrelated to the target site. The onus is on the researcher to execute attended performance testing against limited URLs to ensure HTTP *get* requests per unit time are below an acceptable threshold and returned data volumes per unit time are of acceptable size. In making this decision, the research needs to consider the number of potential site users at peak times and ensure that content download requests will not exceed average peak site traffic levels. Where social networking sites provide developer tools for data access (known as an application programming interface, API) this is less of an issue and a social networking site's API will throttle request rates and downloaded data volumes.

It is probable that institutional review boards will not provide guidelines for researchers to follow when performing web scraping. Informal etiquette guides are published on the internet and researchers should follow good practice. Effectively this is a voluntary code and hence difficult to enforce. The principles of this code are:

- Use download delays to ensure requests do not exceed the number of requests the web site may receive during average peak usage.
- Execute scraping during times when site traffic is at a minimum.
- Google scrapes public web content and caches this content, consider using Google's cache of a website rather than scraping content directly.
- Always respect websites' robots.txt and robots meta tags by ensuring your chosen scraper supports the robots.txt standard (Koster, 1994).
- Do not circumvent technical measures that the website has in place to limit or prevent content scraping.
- Do not attempt to mask your IP address by using proxy services and ensure the headers in the scrapping application's content request include contact details.

2.3.2 Researchers

Online communities and web site administrators may react negatively to the use of their online space and content by researchers, resulting in actions restricting access to content. The majority of websites will have no specific provision for consumption of user generated content for scientific purposes, hence researchers should follow the website's T&Cs and acceptable use policies for standard users where no separate stipulations are available. Following the letter and the spirit of site policies will minimise conflict with the site owners and negative responses from the online community.

The negative reactions of online communities could engender mistrust of researchers, resulting in a barrier to the uptake of science (Wynne, 1992). Records of grievances raised by online communities are likely to persist on the internet for an indefinite period and recent political surprises indicate that social media rhetoric is at least as influential in sentiment formation as scientific research and expert opinion (K. F. Brown et al., 2012; P. N. Howard & Kollanyi, 2016; Rosner, 2017). Personal observations have indicated that participatory stakeholders do not always differentiate between research groups, government organisations and NGOs, presenting a risk that hostility to a specific research group could have unforeseen impacts, for example refusal to participate in government fish harvest reporting schemes and citizen science projects. Although the above is merely posited, the importance of trust in the fisheries scientific community has been recognised (Glenn et al., 2012).

2.3.3 Human stakeholders and species/ecosystems

A major consideration is how research outputs derived from social media could be used by management bodies to change policy. Research based on social media content has difficulties in reproducibility and transparency, and robust population inferences are impossible unless the target population is social media users. As such, data obtained by these methods requires additional care in analysis, interpretation and presentation.

Stakeholder inclusion in policy formation and management is an important part of modern marine management, for example in marine spatial planning (Pomeroy & Douvere, 2008). Lack of engagement with stakeholders is likely to increase the risk of noncompliance with regulations (Jentoft, 2000; Kaplan & McCay, 2004) which could have management impacts, particularly in adaptive co-management contexts where scientists need to be seen as “trusted knowledge brokers in stakeholder networks” (Armitage et al., 2009). The necessity to anonymise social media data makes it less transparent to external scrutiny and hence may undermine stakeholder participation and trust.

The researcher has an obligation to consider how information on natural fisheries resources are used by existing and new participatory stakeholders. Such knowledge includes the distribution in time and space of fish and bait species. Consideration must also extend to the human aspects of fisher activity, such as favoured harvesting methods. Secrecy traditionally exists among locals who wish to protect resources which hold a social and economic value (Maurstad, 2002; Olsen & Thuen, 2013; Svensson, 2016). This paper’s authors encountered an example where recreational fisheries effort data were aggregated to 1 km^2 to protect fishing locations however, examination of seabed features on a hydrographic map revealed precise fishing locations. In this case it is likely that the researchers did not consider carefully enough the appropriate scale at which to aggregate their data to protect the interests of local fishers.

The research should consider if publication is likely to increase pressure on environments and species, or escalate social conflicts. Certainly Maurstad (2002) argued that there is much potential for generating conflict through unintentional sign posting of fisher knowledge without careful consideration. Mined data from social media content is unlikely to have been published with a consensus from the local community as a whole, even where that content was public.

2.4 Conclusions

Almost all passive research that uses legal and publicly accessible web content will pose few ethical difficulties for ethical review boards when data are correctly anonymised. Hence in conducting text and data mining (TDM), the emphasis is commonly placed on scientists to assess the ethics of their research on a case by case basis and the need for regular reassessment during the research process as social media sites are rapidly evolving. Although material on the ethics of social media research outside of the medical sciences is limited, the Association of Internet Researchers (Association of Internet Researchers, 2017) and the review of Townsend and Wallace (2016) provide some guidance. Ethical review boards should be consulted even where the proposed research would not require detailed ethical review. Any concerns the researcher may have should be discussed and mitigating measures documented and perhaps published following the ethos of systematic review methodology (Pullin & Stewart, 2006).

Although it is established that passive social media research does not involve human subjects according to the medical sciences definition (Beauchamp & Childress, 2001; Walther, 2002; World Medical Association, 2001), ecological researchers engaging in social media research should use beneficence and nonmaleficience as the guiding principles in assessing effects on all stakeholders. Page (2012) introduces the utility and application of these principles and suggests methods for their empirical assessment. It may be argued (with some dispute) that “do no harm” has the highest priority and is the least ambiguous, as evidenced by being the primary principle of the Hippocratic Oath (C. M. Smith, 2005).

In applying TDM in fisheries research, it is apparent that no direct harm can come to content authors and individuals referred to in the content when their identification is impossible. Hence correct anonymisation of direct and indirect identifiers (a non-trivial task) (Zhou, Pei, & Luk, 2008) is of the utmost importance. Anonymisation is particularly important when excerpts of data will be published verbatim (e.g. in case studies). Anonymisation will also mitigate against reputational damage to researchers and the owners of social media web sites. Furthermore, following good etiquette in TDM will protect researchers and their associated institutions from reputational damage while guarding mined websites against performance degradation.

Legally, researchers and their funders are assured that using public data for research purposes is protected by derogations in copyright law (but national laws vary). Researchers should ensure TDM is allowed under fair use (or equivalent) copyright legislation in the country in which they are based and in the country where the controlling company of the host site is registered. Additionally, the privacy policies of websites are unenforceable under criminal law, assuming the research does not contravene privacy, data protection or other human rights laws. Researchers must not attempt to circumvent technical measures taken by a website owner to curtail TDM as this could be a criminal offence.

People and communities are an intrinsic part of social media research and to the management of fisheries and related ecosystems. Both the science which underpins management decisions and the management decisions themselves must be equitable, hence the science must be transparent and easily defendable in the public arena. When a researcher uses social media to deliver research outcomes they must ask themselves how well social media scraping fits these criteria, in particular when securing informed consent may be impossible.

Utilization of social media content in ecological research is expected to increase, following the trend of other disciplines. Increased smartphone adoption and internet connectivity in the developing world (International Telecommunications Union, 2014) means that social media could become increasingly useful in understanding small scale and artisanal fisheries where fisheries independent data are unavailable. Nevertheless the absence of informed consent presents difficulties for researchers where management may be affected, particularly accepting the serious limitations of the source data in making population inferences.

Finally, it is important to recognise that the authors' of this paper are not legal professionals and this paper should not be construed as definitive legal advice on this issue or used in such a capacity. Ultimately, compliance with all the requirements of institutional bodies and ethical review boards are of primary importance for the indemnification of researchers against any legal action. The use of social media in fisheries and other areas of science is a highly current and developing area of research and as such this paper draws attention to important ethical considerations when considering the use of social media in research.

Chapter 3

Text and Data Mining of Social Media to Map Wildlife Recreation Activity

Chapter 3 was published in journal *Biological Conservation*.

Monkman, G.G., Hyder, K., and Kaiser, M.J., Text and Data Mining of Social Media to Map Wildlife Recreation Activity. *Biological Conservation*, 228, 89–99 (2018).

KH and MK critically reviewed and revised the article. All other work was that of GGM.

Highlights

- Text and data mining of social media can be used to derive data on the temporal and spatial distribution of wildlife recreation.
- Text and data mining does not require specialist knowledge of natural language processing and machine learning to extract useful information.
- A permutative approach to spatial correlation and inter-rater agreement tests overcomes statistical problems when applying a focal operation to GIS layers.
- Social media mined data can provide more data than directed surveys and expert knowledge where wildlife recreation activity is rare or survey difficulties exist.
- Prosecution of the European seabass by recreational fishers in Wales, peaks in May, June and August and is associated with population centres and good road access.
- Activity patterns met expectations, temporal activity patterns correlated with a national survey assessments, but unknowable biases exist.

3 Text and Data Mining of Social Media to Map Wildlife Recreation Activity

3.1 ABSTRACT

Mining of social media has been shown to be a useful tool for social and biological research (e.g. tracking disease outbreaks). This article outlines an accessible approach to the use of text and data mining (TDM) of social media to gather information on wildlife recreation activity. The spatiotemporal distribution of the shore-based recreational European sea bass (*Dicentrarchus labrax*) fishery in Wales is used as an example. Public online user generated content was mined using automated scraping. Data on fisher activity and fish sizes were extracted and then georeferenced by matching place names to a custom compiled gazetteer. Numbers of trips and spatiotemporal trends in the distribution of activity and catches were estimated. Sea bass prosecution was higher in summer than winter, and gear use and trip durations were consistent during the period 2002-13. Comparisons of TDM with existing surveys showed higher levels of activity and catch, and shorter mean trip durations were estimated using TDM. Monthly activity correlated closely with existing survey data. Spatial and temporal data agreed qualitatively with expert knowledge. This article showed that TDM can be used to describe a wildlife recreation activity, but use of TDM to derive unbiased population level estimates is challenging and more work is required to develop appropriate methods to correct for biases. These methods required no expertise in natural language processing or machine learning; a working knowledge of programming (e.g. in Python or R) is all that is needed to apply this approach. The opportunities to use TDM will increase with the continuing adoption of smartphones in emerging economies and developing nations and may be of particular utility where other data are unavailable.

3.2 INTRODUCTION

Wildlife recreation can bring economic benefits (Hudson & Lee, 2010; Mbaiwa, 2003; Naidoo et al., 2016; Scheyvens, 1999; Wilson & Tisdell, 2003) and participation has been associated with improved health and well-being (A. Brown et al., 2012; Curtin, 2009; Freudenberg & Arlinghaus, 2010; Gartner et al., 2002). Nonetheless, wildlife recreation that involves direct removal of individuals (termed 'consumptive wildlife recreation', Freese, 2012) has been associated with negative effects on populations and environments (Green & Giese, 2004; Lewin et al., 2006; McPhee et al., 2002; Schroeder & Love, 2002; Westera et al., 2003). Understanding the sustainability of such activities necessitates the assessment and monitoring of the target species, the environment, and the recreational activity which exploits those species. Monitoring these activities presents considerable challenges given the often unregulated nature of recreational activities associated with wildlife.

There is increasing interest in using social media data for ecological monitoring and surveillance (Daume et al., 2014; Environ et al., 2010), to improve understanding of participants' practices and behaviour (Belhabib et al., 2016; Hinsley, Lee, Harrison, & Roberts, 2016; D.R. Martin et al., 2012; Richards & Friess, 2015; Shiffman et al., 2017) and the magnitude of participatory activity (Dustin R Martin et al., 2014; Wood, Guerry, Silver, & Lacayo, 2013). Researchers tend to mine data from social networking sites with Facebook being a popular source (e.g. Di Camillo et al., 2018; Eid and Handal, 2017; Mori et al., 2018) however, financial costs are associated with access to historical content and users are not anonymous. In addition, correctly classifying relevant data can necessitate computationally expensive processing (Daume et al., 2014). Discussion forums provide an alternative source of user generated content which can be information rich cf. the public feeds of social networking sites. Wildlife recreation-centric forums are commonplace and researchers have manually extracted data from forums to gain insights into the—sometimes illegal—practices of recreational shark fishers (Shiffman et al., 2017) and Martin et al. (2014) correlated monthly post counts containing named water bodies (n = 19) with monthly effort data from a roving survey.

Recreational fishing has an estimated 1 billion participants worldwide (Cooke & Cowx, 2004) and around 9 million participate in sea fishing in Europe (Hyder et al., 2018; Z. Radford et al., 2018). It is inevitable that any recreational fishing activity will result in significant species removal (Z. Radford et al., 2018) as release rates are rarely 100% (Ferter et al., 2013) and some post-release mortality in a recreational fishery is unavoidable (A. Bartholomew &

Bohnsack, 2005; Lewin et al., 2018). Recreational fishers tend to target specific species based on the fish attributes such as size, fighting prowess and palatability which may explain why the European sea bass (*Dicentrarchus labrax*, henceforth sea bass) is the most sought after fish species of marine recreational fishers (MRF) in several EU countries (Armstrong et al., 2013a; Goudge, Morris, & Sharp, 2010; Herfaut et al., 2012; Hyder et al., 2018; Monkman et al., 2015; van der Hammen, de Graaf, & Lyle, 2016). Pressure on sea bass stocks has grown over the last decade with increasing commercial landings (ICES, 2016b) and a significant harvest by recreational fishers (Armstrong et al., 2013a; Herfaut et al., 2010; Hyder et al., 2018). In 2014, ICES estimated that spawning biomass had approached the limit reference point (ICES, 2014a) hence harvest controls were applied to recreational and commercial sectors from 2015 (European Commission, 2015b, 2015a, 2016b) and are still in place (European Commission, 2018). ICES still considers the stock to be data limited (ICES, 2017a, 2017b).

Sea bass are a particularly important species for recreational and commercial fishers in Wales, UK (Monkman et al., 2015; SEAFISH, 2016), yet there is little temporal or spatial data on the distribution of recreational sea bass fishing activities across Wales. Surveys of marine recreational fishers (MRFs) are problematic because sea bass captures by MRFs are comparatively rare across the MRF population (Armstrong et al., 2013a; Goudge et al., 2010; McMinn, 2013; Monkman et al., 2015). Nevertheless, current legislation requires EU member states to report the amount of fishing mortality that is attributable to MRF-related activity (European Commission, 1992, 2008; UK Parliament, 2009).

This study aims to introduce conservation researchers to the potential for text and data mining (TDM) of social media to provide meaningful information on wildlife recreation recorded by participants online. This article extends the work of Martin et al. (2014) and Shiffman et al. (2017) by describing how stages of the TDM process can be automated to improve data yields per unit time. These methods require no expertise in linguistics, natural language processing or machine learning. It is shown how large volumes of open text can be automatically parsed sentence-by-sentence to extract quantitative descriptors from a difficult-to-survey stratification of people who engage in wildlife recreation. We use the shore-based recreational sea bass fishery in Wales, United Kingdom as a methodological case study to address the following questions.

i. Can text and data mining of social media produce temporal population estimators of key activity metrics, including gear type, fishing platform, gear numbers, trip duration and capture events?

ii. Does forum reported activity provide a reliable proxy of the observed inter-monthly variation in actual activity, which agrees with directed survey results, and expectation?

iii. Can indicators of the spatio-temporal distribution of an important wildlife recreation activity be derived from TDM of posts to discussion forums?

iv. What are the general limitations in using social media data in wildlife research?

3.3 METHODS

3.3.1 STUDY AREA

The study area was recreational shore fishing activity which resulted in the capture of sea bass from the inshore territorial waters of Wales, UK (Figure 3-1). Henceforth the term sea bass prosecution will be used to mean target, catch and harvest of sea bass. All references will refer to sea bass prosecution from the shore, unless otherwise stated. This definition of recreational fishing accords with that of Pawson (2008), noting that rod-and-line fishing (angling) was the only method encountered on discussion forums. The Welsh shoreline is over 2740 km in length and is highly varied. Much of the coastline could be visited by MRFs who almost exclusively use rod-and-line gears for fish capture (GGM personal observations).

3.3.2 ETHICS

Only posts made to public web forums were mined. The name and unique resource locator (URL) of the forums used are not disclosed within this study to protect the identity of forum operators. Some forum users can become hostile to researchers (Nardi, 2015) and this could negatively impact the forum. Forums whose terms and conditions restricted access by automated web crawlers or where such access was blocked within a robots.txt file were excluded. No personal details of forum members, including pseudonyms, were persisted to permanent disk storage. Data were stored in an encrypted Microsoft SQL Server database and all downloaded raw text was deleted after processing. Although specific locations of activity were identified, all such data were aggregated to regions of at least 50 km^2 and temporal data were aggregated within months.

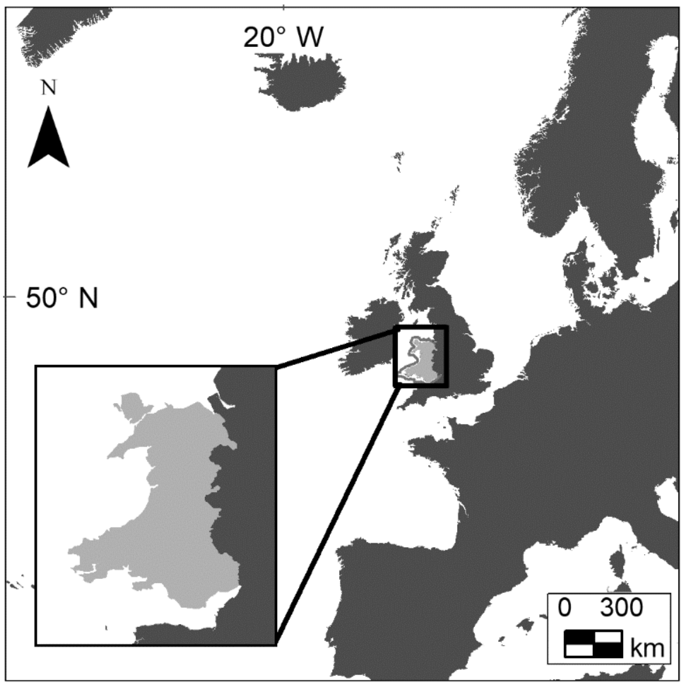

Figure 3-1. Geographical scope, Wales (inset), United Kingdom.

3.3.3 Social Media Text and Data Mining

3.3.3.1 Discussion forum identification

Google World Wide Web searches were executed during September 2013 to identify discussion forums with MRF content. The first search term was *(Wales OR Welsh) (angling OR fishing) (sea OR marine) (forum OR newsgroup OR board OR bulletin OR usenet) site:.uk*. The returned sites were reviewed in descending order of relevance (according to Google's PageRank). If a site had 10 or more catch report threads submitted per month to any board, in any month up to Sept 2013 then the URL was recorded for later use. Once no new site had been identified after 60 minutes of searching and assessing sites, the process was repeated with a 30 minute search without the site:.uk search directive.

3.3.3.2 Crawling and scraping

The rules used by each forum to generate opening post URLs were determined. Figure 3-2 illustrates the hierarchical structure common to online discussion forums. Using these rules, a list of thread URLs was programmatically generated (over 100 000 distinct URLs in total) for each relevant board across

the eight forums. The opening post of each thread contained the target content because the opening post typically reports a single fishing trip. Typically an opening post has multiple follow-up responses which rarely detail additional trips (GGM personal observation), making follow up responses computationally expensive to process for little return. The title, post date and text were all scraped from the opening post using Outwit Hub (Outwit Technologies, 2013). Threads within a board have the same document markup structure, hence eight separate scrapers were created,

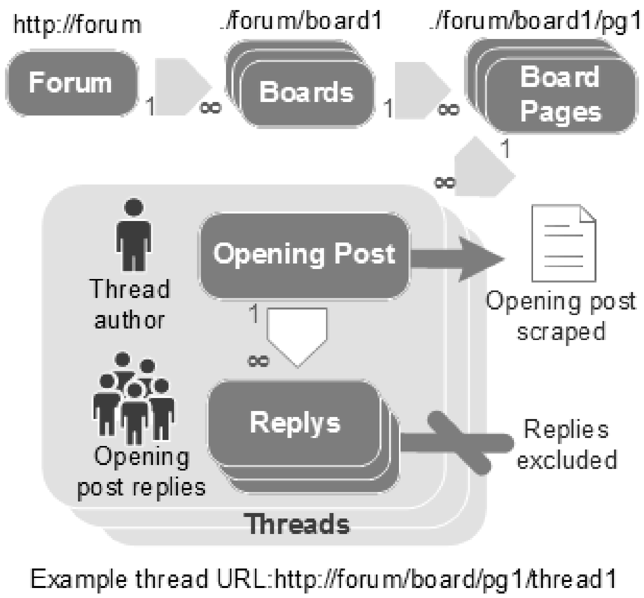

Figure 3-2. Hierarchical structure of a typical world wide web discussion forum.

Symbols 1 and ∞ show a one-to-many relationship. Hence a forum has many subject boards, which has many pages containing many threads. A thread is created with an opening post. Only the content of the opening post was mined.

one for each forum. For all eight boards it was sufficient to define unique enclosing markup tags (e.g. *<title></title>*) to extract the required content. Scraped data were stored in an encrypted Microsoft SQL Server database (Microsoft, 2008) for later processing. Forums were dominated by users in the United Kingdom hence all scraping was executed between 23:00 and 07:00 UTC to ensure web servers were not put under undue stress. A general introduction to TDM is provided in online Appendix A.

3.3.3.3 CLASSIFICATION, INFORMATION EXTRACTION AND GEOREFERENCING

A lexicon of common terms used on social media by MRFs were compiled using several different approaches as outlined in Appendix D. These were primarily from unstructured face to face interviews with three MRFs who fish for sea bass, from the authors' own experience and review of a sample of forum posts with bass catches. The lexicon included spelling errors, colloquial terms and plural word forms. Synonyms used for sea bass (e.g. silver bullet), fishing platform (e.g. boat), activity durations, date-times, weight, length and scalar measures (e.g. couple), geographic locations, tidal state (e.g. high water) and gears (e.g. rod) were collated. Figure 3-3 summarises the number of terms (circled) used according to their lexical classification.

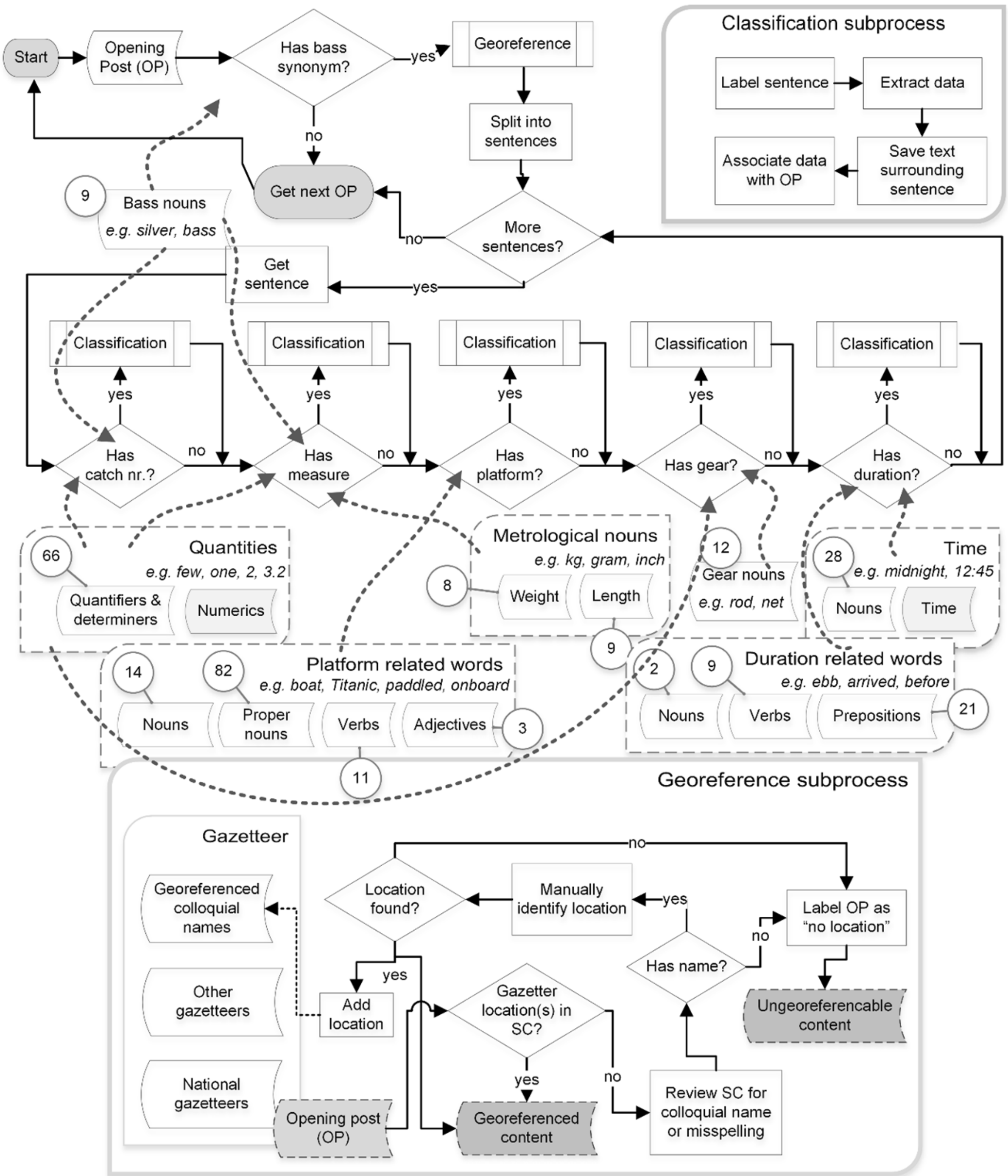

Figure 3-3. Text and data mining steps. Opening posts (OP) were scraped, then processed if they contained a synonym of European sea bass. The OP was split into sentences and each sentence classified according to its content. Dashed boxes and dashed connectors show the lexicon used to identify relevant content. The circular call-out labels (e.g. 66 in box Quantities) give the number of words or phrases in the lexicon, noting that plural word forms, although used, are not included in the counts. SC = scraped post.

A custom gazetteer of place names with their spatial coordinates was compiled from multiple sources. The bulk of entries came from the Seazone and Ordnance Survey gazetteers (Seazone 2014, Ordnance Survey 2015) from which all named features within 6 nautical miles seaward and landward of the mean high water line were included. Welsh place names and location colloquialisms used by MRFs were also added. To protect participants' fishing locations and identity in the stored data, all locations were assigned to the nearest landward settlement in Google Maps then snapped to the mean high water spring tide line. Spatial processing was carried out in ArcMap 10 (ESRI, 2010).

Scraped text was processed with a bespoke software application according to Figure 3-3. Only opening posts containing at least one sea bass synonym were processed. Opening posts were split into sentences using the SharpNLP (SDragon, 2006) implementation of OpenNLP (Hornik, 2016). Sentences were checked against the lexicon and labelled in the database according to the sentence content (e.g. a sentence containing the word "rod" would be labelled as "gear"). Multiple labels were allowed, for example, a sentence containing the gear keywords "rod" and "net" would be labelled with both terms so that the sentence could be manually reviewed to determine if usage referred to rod-and-line fishing (angling) or a different fishing method.

Sentences which contained a sea bass synonym, and a numeric (2, 2.5) or a quantifier or determiner (e.g. one, couple) were labelled as having a catch number. Sentences with a metrological noun (e.g. kg, pound) were classified as containing a measure of length or weight (according to the noun). Regular expressions were used to determine if a sentence had a time or a numeric, the detection routines are in Appendix B. Simple keyword matches were used to classify the sentence as containing gear. In addition to standard words indicating the platform activity, the lexicon included proper nouns of for-hire boat names and the manufacturer and model names of boats and kayaks.

Following filtering and classification, the number of pertinent sentences was relatively small, hence the development of an automatic classifier was unjustified. Quantifiers were converted to numerics (e.g. couple to 2) and then numerics were parsed from sentences and saved with the keyword label associated with the sentence. The surrounding text was also recorded to provide context during manual review. All extracted sentences were then manually reviewed and corrections made to the parsed measures as necessary. Forum users could contribute to more than one social media site, because some sites allow pseudo-anonymity then the same trip could be counted multiple times. Duplicates were easily identified by ordering the data by date, location and the website source.

During the manual review process, the quality of the activity parameters were scored between 1 and 5 (henceforth quality rank) with 1 indicating that substantial interpretation from the author's personal experience was required for gear number, fisher number and duration, to 5 indicating that the three parameters were all unambiguously expressed in the report. All activity related calculations excluded records with a quality rank < 3. Examples of a quality rank of 3 include some ambiguity in duration (e.g. fished from ½ tide to around low) and an assumption of single gear use based on the contents of the post (e.g. lure fishing necessitates the use of a single rod).

Sea bass quantity and individual metrics (e.g. length and weight) were excluded from a detailed analysis here due to the sensitivities of the online communities. MRF activity was expressed as gear hours per trip, i.e. the product of trip duration, gear number and fisher number.

3.3.4 General statistical methods

For parametric tests, normality was determined using a Shapiro-Wilk (Royston estimation) test. Homoscedasticity was validated with Bartlett's test for parametric data. Where assumptions were not met for parametric tests, a 1000 sample bias corrected accelerated bootstrap (BCa) was used for estimates of central tendency and Kendall's tau-b for correlation testing. Bootstraps were executed in IBM's SPSS 20 (IBM Corp, 2011). All other analyses were performed in R (R Core Team, 2016) except where otherwise stated. An α of 0.05 was used for all significance tests. For time series data, trends were tested for using the Hamed-Rao modified Mann-Kendall (HRMK) method (package Addinsoft, 2014). Differences between factor levels following a significant parametric ANOVA were determined using the Tukey-Kramer post-hoc test because groups were mildly unbalanced (Hayter, 1984).

3.3.5 Monthly activity patterns

Monthly patterns of sea bass prosecution were investigated over the period where participants published trip events to social media (March 2002 – September 2013). Calculation details and equations for named variables are given in sections 3.3.5 and 3.3.6, and in Appendix C. The following measures of monthly activity were calculated and are reported in the results.

(i) $\{M\}_{|140|}$[2], time series of unadjusted *monthly activity* (gear hours) between March 2002 and September 2013.

[2] The vertical bar notation (e.g. |140|) is the number of elements (cardinality) of an array, **not** the modulus function.

(ii) $\{Z\}_{|140|}$, time series of *standard monthly activity*. The monthly activity $\{M\}$ differenced by the mean monthly activity for a year, divided by the standard deviation of mean monthly activity for the year.

(iii) $\{\acute{Z}\}_{|140|}$, time series of *y-j standard monthly activity*. The Yeo-Johnson transform of $\{Z\}$, homoscedastic.

Mined activity was validated against trip durations collected during the single year of the Sea Angling 2012 on-site roving creel survey of shore fishers (abbr. SA2012 survey; CEFAS, 2012; Armstrong *et al.*, 2013) where a recorded trip had at least one catch of an Atlantic cod (*Gadus morhua*, henceforth cod) or sea bass. The following variables are reported.

(i) $\bar{\acute{Z}}_{|12|}$, *mean y-j standard monthly activity* from the time series $\{\acute{Z}\}$.

(ii) $\bar{Z}_{sa_{cod}|12|}$, standardized (Z) scores of total monthly trip hours from SA2012 survey trips with a recorded cod capture.

(iii) $\bar{Z}_{sa_{bass}|12|}$, as $\bar{Z}_{sa_{cod}}$, for trips with a recorded sea bass capture.

Cod was chosen as an additional validation dataset to support the assumption that the distribution of $\bar{Z}_{sa_{bass}}$ was not an artefact or chance occurrence arising from the SA2012 survey process. The prior expectation was that cod prosecution from the shore has a maximum during winter and a minimum in summer. No adjustment was made for monthly sampling effort because sampling effort was a poor predictor of total monthly prosecuting trip duration (BCa bootstrap GLM, $F_{(1,34)} = 0.07$, $p = 0.62$, $b = -0.002$, $R^2 = 0.002$).

3.3.6 Seasonal and spatial activity patterns

The spatial and seasonal distribution of effort for the coastline of Wales was mapped between May and October (summer) and November to April (winter). Twenty four regions were created using a 25 km by 25 km grid, intersected with the 6 nautical mile national limit and the smoothed mean high water line. Polygons under ~50 km^2 were merged with neighbouring regions which shared the marine area characteristics according to Parker (2015). All calculations were derived from the trended time series $\{M\}$ described in 0. The following variables are reported.

(i) $\bar{\boldsymbol{x}}_{\boldsymbol{season}\,|2|}$ *Mean seasonal activity* (gear hours region^{-1} year^{-1}), unadjusted mean of gear hours for season aggregated for all regions.

(ii) $\mathbf{I}_{\boldsymbol{sr}\,|24|}$, *Intensity* (gear hours km^{-1} month^{-1}), standardised by coastline length, stratified by summer and winter across the 24 regions to give a total of 48 strata.

(iii) $\mathbf{D}_{sr\,|24|}$, *Differenced intensity* (gear hours km^{-1} $month^{-1}$), $\mathbf{I}_{sr}$ differenced using yearly intensity means.

3.4 RESULTS

From social media, eight UK centric forums had threads detailing sea bass prosecution in Wales from 2002. Increased forum traffic was observed into the mid-2000s, peaking at ~9800 threads (of any type) across the eight forums in 2009. Forums yielded 20 060 threads containing a sea bass noun synonym and these threads were downloaded over approximately 56 hours of processing time. Of the 20 060 threads, 14 853 (74%) were successfully georeferenced from the compiled gazetteer. Of these 14 853 threads, 4040 individual sentences from 1110 (6%) threads matched keywords allowing the extraction of a spatially referenced sea bass prosecution event. Of these 1110 threads, 973 (88%) referred to prosecution from the shore. For-hire boat, private boat and kayak prosecution events accounting for 3%, 6% and 4% of the total respectively, there was insufficient data for these platforms to produce any meaningful spatial or temporal analysis. Two forums contributed 90% of all records. A total of 463 catch per unit activity measures were obtained. From 2002 to 2013 a total of 1456 separate sea bass lengths or weights were obtained (annual mean ±SD, 81.3 ±50.5 measures per annum).

3.4.1 TRIP DURATION AND GEAR USE

A total of 2303 shore trip hours with a quality rank >3 were scraped from social media. 363 trips included the parameters fisher number, gear number, trip duration and trip length. Trip duration was subject to pronounced response heaping, with number preferencing or prototyping, in particular at 4 hours, a common heaping digit, see Rodgers et al. (1993).

The mean duration of a trip was 3.9 hours [95% CI 3.7, 4.1] and MRFs used between 1 and 4 rods (median 1, IQR 1–2, M 1.6 [95% CI 1.5, 1.7]). These data equated to mean activity of 6.7 gear hours fisher^{-1} [95% CI 6.1, 7.3] and mean catch per hour of 0.72 sea bass hour^{-1} [95% CI 0.60, 0.84].

The number of gears used by MRFs on a trip did not change significantly between 2003 and 2013 (multinomial logistic regression main effect year, $\chi^2_{(0,\ 16)} = 21.94$, $p = 0.16$). The proportion of single gear use also remained largely unchanged between 2005 and 2013 (linear regression, $F_{(1,7)} = 0.86$, $p = 0.38$, $R^2 = 0.11$) however, the negative regression coefficient (-0.48 single gear proportional use year^{-1}) suggested a weak but increasing trend in gear use per trip between 2005 and 2013. There was a marginally significant increase of 0.23 gear hours trip^{-1} year^{-1} between 2005 and 2013 (BCa bootstrap GLM, $F_{(8,311)} = 1.91$, p = 0 06, R = 0.22). Trip durations were not influenced by the spring, summer, autumn and winter seasons (ANOVA,

$F_{(3,316)} = 0.43$, $p = 0.73$). The number of gears used on a trip had no significant influence on sea bass catches (BCa bootstrap GLM, $F_{(2,336)} = 1.85$, $p = 0.16$, $b = 0.18$ sea bass catches gear^{-1} hour^{-1}, [95% CI -0.16, 0.41]) despite the quoted increasing trend.

3.4.2 MONTHLY ACTIVITY PATTERNS

There was an increasing trend in forum reported activity $\{M\}$ between 2003 and 2013 (Figure 3-4b, Mann-Kendall Hamed & Rao method; Gear activity: $\tau_{(139)} = 0.46$, $p < 0.001$; Trips: $\tau_{(139)} = 0.67$, $p < 0.001$). Standard monthly activity $\{Z\}$ showed strong seasonality and tended to follow increasing or decreasing cycles, as opposed to untrended monthly fluctuations (Figure 3-4a & b). Yeo-Johnson standard monthly activity $\{\acute{Z}\}$ had a pronounced positive autocorrelation coefficient (ACC) at a one month lag (Figure 3-4c, $ACC = 0.48_{(n = 139)}$, [95% CIs -0.16, 0.16], $p < 0.001$) and significant positive autocorrelations between 10 and 14 months. The significant negative correlations between 4 and 8 months shows that high activity tends to follow low activity in 6 monthly cycles. Overall, the magnitude of activity for a given month between years was stable, with the autocorrelation coefficient maximum occurring at a lag of 12 months ($ACC_{(n = 139)} = 0.59$, 95% [CIs -0.16, 0.16], $p < 0.001$).

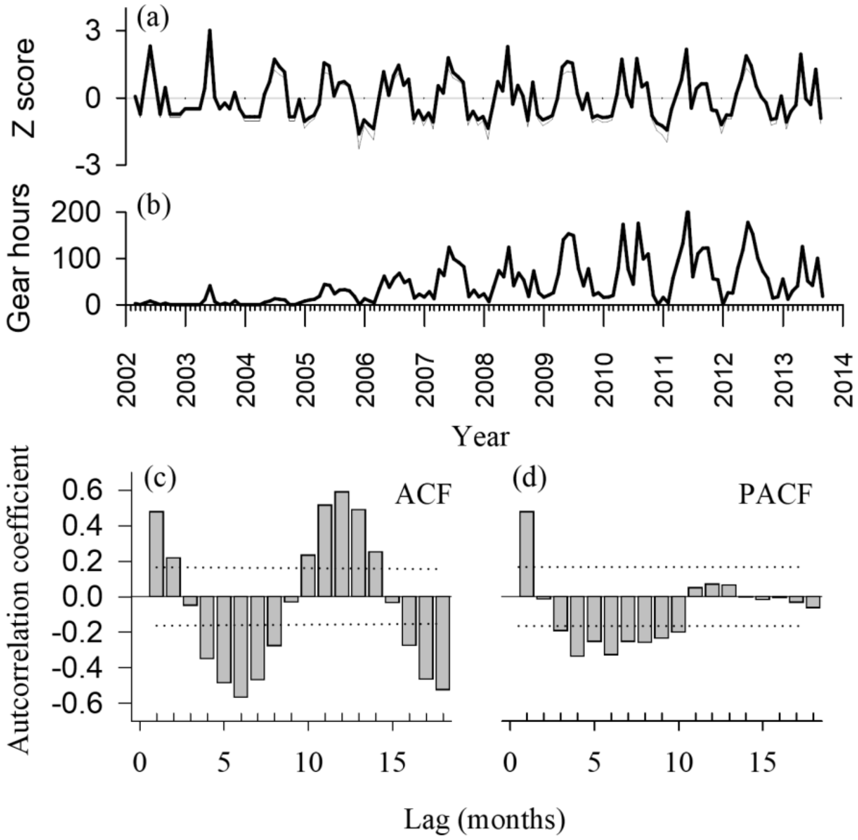

Figure 3-4. (a) Time series of standard monthly activity $\{Z\}$, aggregated from discussion forum mined trips where European sea bass were prosecuted by recreational fishers on the coastline of Wales between March 2002 and September 2013. The faint line is the Yeo-Johnson transformation of $\{Z\}$ ($\{\acute{Z}\}$, $\lambda = 0.4$), noting the transform is mild. (b) Monthly activity ($\{M\}$, gear hours), the unadjusted monthly sum of activity in gear hours. (c) and (d) shows the autocorrelation (ACF) and partial autocorrelation (PACF) structure of time series (a) with 95% confidence intervals (dotted line) around the mean of 0. Autocorrelation coefficients lying outside the confidence intervals are significant at $\alpha = 0.05$. Bracketed letters indicate time series labels as described in online Appendix C.

Activity peaked in June, May and August (Figure 3-4a; Mean ±SD: June, 1.70 ±1.00; May, 0.93 ±0.61; August, 0.66 ±0.76) and several years had a pronounced bimodal activity peak over the summer months, with the summer minimum typically occurring in July, this contributes to the negative cross correlation observed at lag 3 in Figure 3-4d. The least fishing activity tended to occur in February (-0.96 ±0.29), January (-1.24 ±0.37) and December (-0.78 ±0.40). Standardized mean monthly activity $\bar{\bar{Z}}_m$ differed significantly between months and the effect of month on the level of activity was large (Figure 3-5, ANOVA, $F_{(11, 127)} = 19.9$, $p < 0.001$, $\omega = 0.77$). The months of May through to September all had activity exceeding the standard normal mean of 0 and the split of summer high activity months and winter high activity months was significant (Tukey-Kramer post-hoc, Figure 3-5).

The observed monthly trends in sea bass prosecution agreed broadly with those derived from the SA2012 on-site survey (Armstrong et al., 2013a; CEFAS, 2012), particularly in the winter months of November to March when relative activity was low (Figure 3-5). This agreement is supported by the cross-correlation structure between $\bar{Z}_{sa_{bass}}$ and $\bar{\bar{Z}}$, with significant cross correlations only occurring at a lag of 1 month, and marginally at a lag of 0 months (cross correlation, 0 month lag, CCC(n = 12) = 0.54, [95% CIs -0.57, 0.57], p = 0.07; 1 month lag, CCC(n = 12) = 0.62, 95% [CIs -0.59, 0.59], p = 0.05), noting that the cross correlation function has a normal distribution with a mean of 0. A significant negative cross correlation occurred when comparing $\bar{Z}_{sa_{cod}}$ and $\bar{\bar{Z}}$ at lags of 0 and -1 (cross correlation, 0 month lag, CCC(n = 12) = -0.654, 95% CIs [-0.57, 0.57], p = 0.03; -1 month lag, CCC(n = 12) = -0.67, 95% CIs [-0.59, 0.59], p = 0.03). As expected, activity resulting in cod captures was elevated in the winter months of November to March (Figure 3-5).

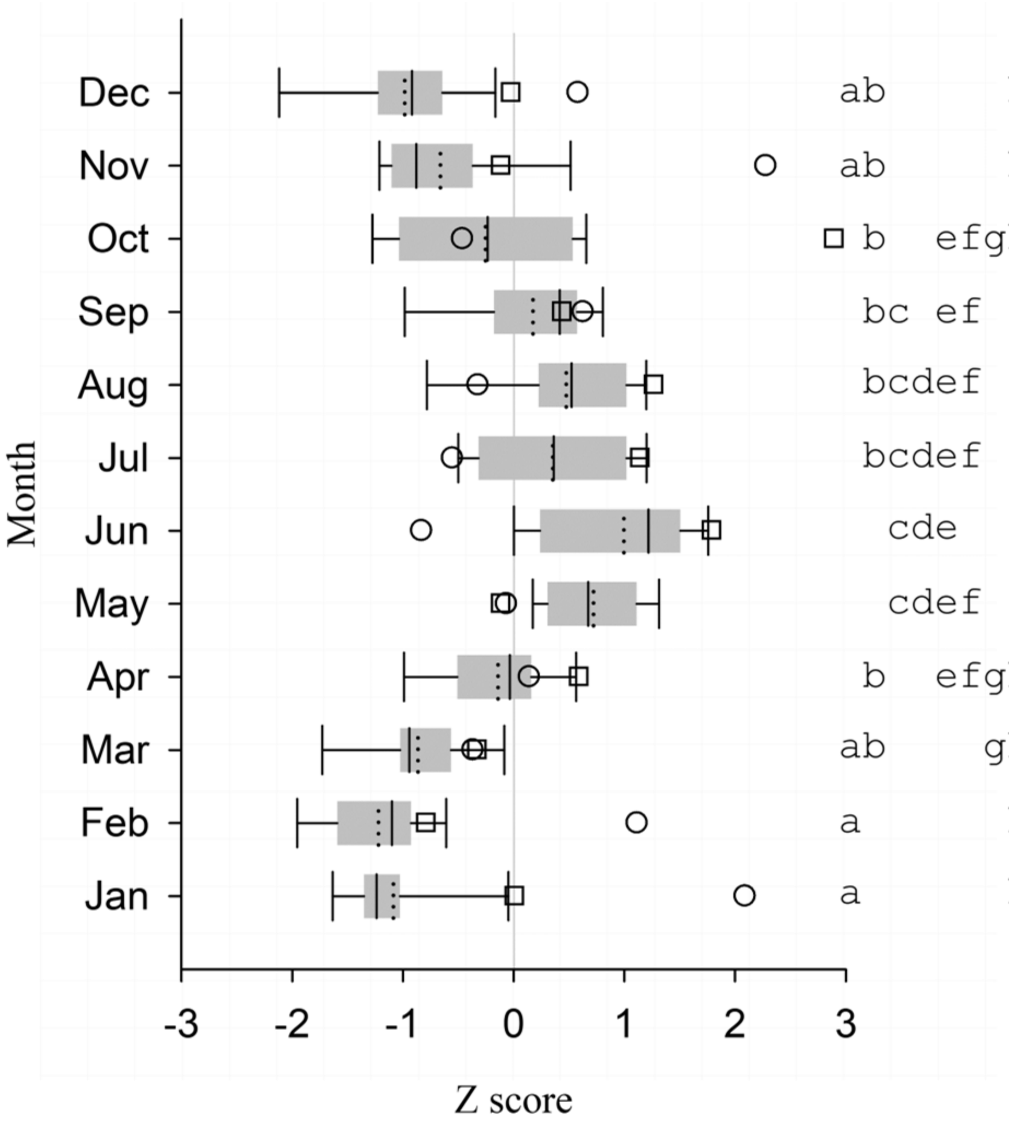

Figure 3-5. Box plot of mean Yeo-Johnson standardized monthly activity ($\bar{\bar{Z}}$), aggregated from discussion forum data where European sea bass catches were reported by recreational fishers from the coastline of Wales between March 2002 to September 2013. The within-box dotted line is the mean, dashed line is the median. Letters show non-significant groups from post-hoc pairwise comparisons (Tukey-Kramer, $\alpha = 0.05$). Standardised monthly mean activity (variables $\bar{Z}_{sa_{cod}}$ and $\bar{Z}_{sa_{bass}}$) of shore trips with recorded catches of sea bass (□) and cod (○) from the Sea Angling 2012 on-site survey are overlayed (CEFAS, 2012).

3.4.3 Seasonal and spatial activity patterns

Mean seasonal activity (x_{season}, gear hours region^{-1} year^{-1}) for summer was 3 times that of winter (mean ±SD summer = 21.6 ±31.0; winter = 7.3 ±11.2; BCa bootstrap ANOVA, $F_{(1, 382)} = 36.6$, $p < 0.001$). Location had the greatest influence on differenced intensity, D_{sr}, ($p < 0.001$, $\omega = 0.56$), but season ($p < 0.001$, $\omega = 0.29$) and the interaction of season and location ($p < 0.001$, $\omega = 0.31$) were both significant in their effect on intensity (Table 3-1). North east Wales had the highest intensity (Figure 3-6, Table 3-2), and the maximum country wide intensity recorded in region NE2 (Mean ±SD, 0.25 ±0.18), which was ~200% greater than the next highest intensity recorded in region NW8 (Mean ±SD, 0.11 ±0.08). Intensity above the annual mean was also observed during summer in south east Wales (regions S2, S3, S4, S5 and S6).

Table 3-1. Results of bias corrected accelerated bootstrap general linear model on the intensity ($\mathbf{I}_{sr}$, gear hours km^{-1} month^{-1}) for season (Summer, May – October; Winter, November – April) and for locations as shown in Figure 3-6. The effect size ω can be interpreted as medium (++) when ω ⩾ 0.3 and large (+++) when ω ⩾ 0.5. Model R^2 = 0.56.

Factor (fixed)	**df**	**F**	**P**	**ω**
Location	23, 336	11.2	< 0.001	0.56^{+++}
Season (Summer or Winter)	1, 336	64.7	< 0001	0.29^{++}
Location x Season	23, 336	4.2	< 0.001	0.31^{++}

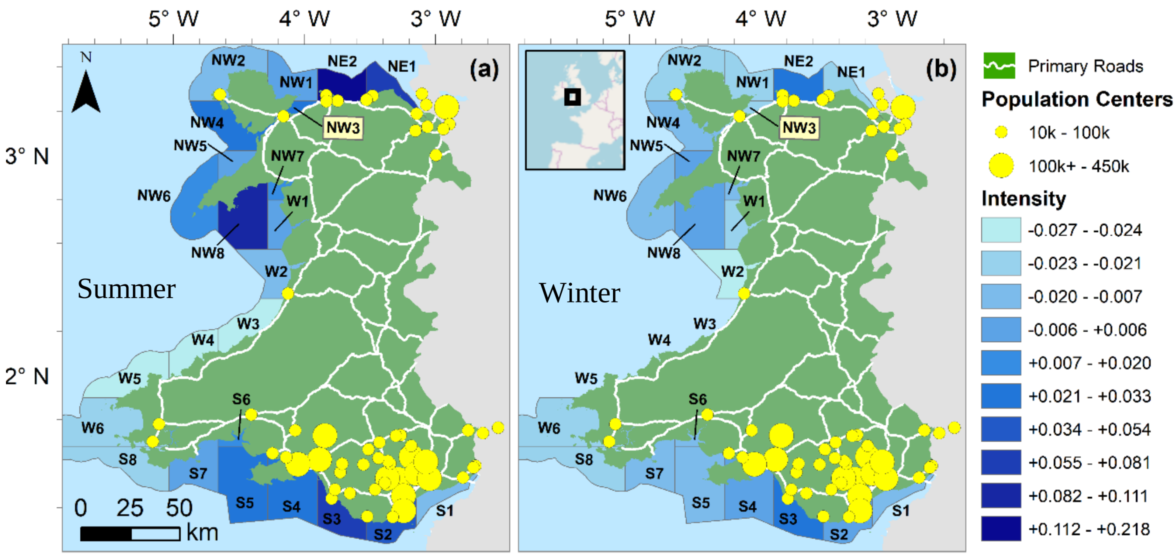

Figure 3-6. Differenced intensity ($\mathbf{D}_{sr}$, $x - annual\ mean\ intensity$, gear hours km^{-1} month^{-1}) for winter (November – April) and summer (May – October). Classes assigned using a 10 group Jenks classifier. Intensity was derived from text mining discussion forums where European sea bass catches were posted by recreational fishers from January 2006 to September 2013 around the coast of Wales, UK.

Regional intensity levels tended to be preserved across seasons, i.e. popular summer regions also tended to be popular in winter and vice versa (Kendall's tau-b, $\tau(24) = 0.64$, $p < 0.001$). Winter intensity never exceeded summer intensity for any region, with significant interactions of season*region (Table 3-1, p < 0001) attributed to greater proportional decreases in intensity at locations with high summer intensity, in particular in north east Wales (NE1, NE2) and NW8 (southern Llŷn Peninsula). Summer and winter intensity (gear hours km^{-1} month^{-1}) varied across regions by up to 2 and 1 orders of magnitude respectively (summer range 0.001 – 0.246; winter range 0.002 – 0.062). This was supported by the simple effects comparison, with season having a medium effect at NE2 ($R = 0.44$), and small effects at NE2 NW3, NW4, NW8, S2 and S4 (Table 3-2, $0.1 \leqslant R < 0.3$).

Table 3-2. Intensity ($\mathbf{I}_{sr}$) and differenced intensity ($\mathbf{D}_{sr}$, $x - annual\ mean\ intensity$) per kilometre of smoothed high water coastline per month (gear hours km^{-1} month^{-1}) derived from trip records of forum reported European sea bass catches made from the coastline of Wales between 2006 and 2013. NR = No recorded activity. F, P and R give results of simple effects analysis on the effect of season on differenced intensity across regions. Effect size (R) key: +, ⩾0.1 - <0.3; ++, ⩾0.3 - <0.5. N, n; N is the total trip numbers between 2006 and 2013 recorded in the region, n is the count of annual gear hours km^{-1} month^{-1} means used to calculate the presented figures (i.e. the count of years).

	Summer (May – October)			**Winter (November – April)**			**Simple effects (season*region)**		
Region	**N, n**	**Intensity ±SD**	**Differenced intensity±SD**	**N, n**	**Mean intensity ±SD**	**Differenced intensity±SD**	**$F_{(df=1336)}$**	**P**	**R**
NE1	37, 8	0.110 ±0.098	0.081 ±0.095	3, 8	0.007 ±0.013	-0.022 ±0.020	24.7	<0.001***	0.26^{+}
NE2	146, 8	0.246 ±0.175	0.218 ±0.170	38, 8	0.061 ±0.031	0.033 ±0.028	79.1	<0.001***	0.44^{++}
NW1	18, 8	0.033 ±0.029	0.004 ±0.032	1, 8	0.005 ±0.014	-0.023 ±0.020	1.80	>0.05	0.07
NW2	25, 8	0.014 ±0.011	-0.014 ±0.007	4, 8	0.005 ±0.010	-0.024 ±0.007	0.21	>0.05	0.03
NW3	50, 8	0.053 ±0.033	0.024 ±0.026	7, 8	0.006 ±0.009	-0.023 ±0.013	5.17	<0.05*	0.12^{+}
NW4	61, 8	0.054 ±0.030	0.025 ±0.029	18, 8	0.017 ±0.016	-0.012 ±0.011	3.16	>0.05	0.10^{+}
NW5	13, 8	0.028 ±0.059	0.000 ±0.057	5, 8	0.023 ±0.031	-0.005 ±0.034	0.062	>0.05	0.01
NW6	11, 8	0.044 ±0.053	0.016 ±0.046	4, 8	0.016 ±0.032	-0.013 ±0.033	1.88	>0.05	0.07
NW7	19, 8	0.043 ±0.051	0.015 ±0.047	3, 8	0.008 ±0.012	-0.021 ±0.012	2.86	>0.05	0.09
NW8	51, 8	0.139 ±0.083	0.111 ±0.077	10, 8	0.035 ±0.027	0.006 ±0.021	25.2	<0.001***	0.26^{+}
S1	6, 8	0.016 ±0.022	-0.012 ±0.020	4, 8	0.012 ±0.027	-0.017 ±0.022	0.045	>0.05	0.01
S2	32, 8	0.083 ±0.071	0.054 ±0.073	16, 8	0.034 ±0.021	0.006 ±0.023	5.50	<0.05*	0.13^{+}
S3	64, 8	0.099 ±0.045	0.070 ±0.045	32, 8	0.062 ±0.051	0.033 ±0.054	3.13	>0.05	0.10^{+}
S4	41, 8	0.049 ±0.018	0.020 ±0.014	24, 8	0.021 ±0.017	-0.007 ±0.016	1.71	>0.05	0.07
S5	25, 8	0.051 ±0.041	0.023 ±0.039	9, 8	0.012 ±0.010	-0.016 ±0.010	3.51	>0.05	0.10^{+}
S6	16, 8	0.048 ±0.087	0.020 ±0.083	9, 8	0.014 ±0.016	-0.014 ±0.017	2.70	>0.05	0.09
S7	30, 8	0.027 ±0.018	-0.002 ±0.019	17, 8	0.015 ±0.015	-0.014 ±0.017	0.34	>0.05	0.03
S8	4, 8	0.005 ±0.010	-0.023 ±0.016	3, 8	0.006 ±0.012	-0.022 ±0.018	0.005	>0.05	<0.01
W1	17, 8	0.028 ±0.031	-0.001 ±0.032	6, 8	0.007 ±0.008	-0.021 ±0.013	0.98	>0.05	0.05
W2	10, 8	0.013 ±0.015	-0.015 ±0.011	2, 8	0.002 ±0.006	-0.026 ±0.009	0.29	>0.05	0.03
W3	1, 8	0.001 ±0.004	-0.027 ±0.010	0, 8	NR	NR	0.004	>0.05	<0.01
W4	1, 8	0.001 ±0.003	-0.027 ±0.010	0, 8	NR	NR	0.003	>0.05	<0.01
W5	1, 8	0.001 ±0.004	-0.027 ±0.010	0, 8	NR	NR	0.004	>0.05	<0.01
W6	11, 8	0.007 ±0.007	-0.022 ±0.010	9, 8	0.007 ±0.013	-0.021 ±0.013	0.001	>0.05	<0.01

3.5 DISCUSSION

In the present study it was shown that the partially automated extraction of data from data dense social media sources of open text can provide information on the spatio-temporal patterns of wildlife recreation across a large area. Furthermore, the temporal patterns were validated against a statistically sound national survey (Armstrong et al., 2013a) and agreed with expectations of strong seasonal differences in sea bass prosecution. The months of maximum and minimum prosecution were derived with good (but imperfect) rank agreement with that of Armstrong et al. (2013a). Other metrics were extracted from social media data, including inter-annual trends in gear use and trip duration, and that the number of gears used was not a significant predictor of catch.

Although discussion forum posts are unstructured text, the density of relevant text made the scraping and classification of data viable, without specialist knowledge of natural language processing or machine learning. Conversely, text and data mining (TDM) of Twitter would demand more complex approaches because the Twitter API provides no access to historical data older than 30 days and there is no entity which amalgamates tweets by special interest groups. However, Twitter's model is untypical, with other SNSs (e.g. Facebook and Google+) allowing users to create forum board like entities dedicated to particular interests, such as hunting or fishing which can be used to extract ecological information (e.g. Mori, Di Bari and Coraglia, 2018). Facebook provides API access to their "Groups". TDM of SNS content is ethically more challenging because users are not pseudonymous. SNS entities dedicated to particular content (e.g. Facebook Groups) can have restricted access although this does not preclude its use for appropriate scientific endeavour (Monkman, Kaiser, & Hyder, 2018).

Social media has been used to derive participant preferences in nature-based tourism. The primary approach has been to manually classify the content of geotagged images scrapped from SNSs and use image class frequency as a proxy for the measure of interest for a defined spatial region. Studies include determining tourist site preferences, (Wood et al., 2013) and assigning the value that tourists place on species and biodiversity (Hausmann et al., 2018; Willemen, Cottam, Drakou, & Burgess, 2015). This article presented an alternative, from which we can derive trip durations, gear use (type, number) and biological measures of the prosecuted species and harvest counts. Trip frequencies and species preferences could also have been extracted. The different methodologies should be seen as complimentary and not competing. Text, image and associated metadata are now frequently published in a single social media post and the increasing availability of machine learning APIs will increase opportunities to automatically process social media content to inform wildlife management policy (Di Minin, Fink, Hiippala, & Tenkanen, 2018).

In the absence of other information, TDM of social media could provide baseline data on activity, such as per-trip effort, hunting or harvesting methods and species preferences. This research found no evidence of gear changes between 2002 and 2013, nor any shift from shore to afloat platforms, despite afloat platforms harvesting ~300% more sea bass by weight in 2012 (Armstrong et al., 2013a) . Such apparent inertia should be interpreted with care as social norms and network externalities could make online communities inelastic to changes in the wider population (Wang & Chen, 2012). Perhaps the most pertinent application of collection spatial and temporal information on wildlife recreation is in planning traditional surveys, where such data can add to expert and local knowledge which is usually used in early stage survey design to for example stratify sampling by periods and geographical area of high and low activity.

3.5.1 Comparing Data Volumes

The number of geospatially referenced trips recording sea bass prosecution exceeded those recorded during the Sea Angling 2012 on-site survey for England (Armstrong et al., 2013a), as did the number of sea bass (*Dicentrarchus labrax*) size estimates (463 vs 67 trips and 1456 vs 114 size estimates). Moreover the 2016 NOAA Marine Recreational Information Program (MRIP) catch surveys (National Oceanic and Atmospheric Administration, 2016) yielded 533 capture recordings of the striped bass (*Morone saxatilis*) from recreational fishers. This suggests that large scale on-site surveys used to assess recreational prosecution nationally present statistical challenges when attempting to measure the distribution of activity in a comparatively rare segment of the general population. Other factors may contribute to low sample representation of sea bass fishers during site randomised surveys, e.g. nocturnal fishing and remote location (Armstrong et al., 2013a; K. Jones, 2009). Such behaviour patterns may be typical of other forms of consumptive wildlife recreation where the nature of the activity tends to reduce encounter rates with participants when conducting surveys.

Other participant knowledge sources have been used to investigate marine recreational fishing (MRF) activity. Both Bennett et al. (1994) and Gartside et al. (1999) used MRF club competition records, which provided an extensive time series of catches (>20 years) from which an impressive 15 763 and 35 682 fish measures were derived. Clubs and their events follow a standard set of rules (e.g. gear restrictions) and tend to be held at the same time and location year on year. It is also likely that social factors tend to homogenise activity. This homogenisation and the large number of data points derived allowed the authors to demonstrate that their club derived CPUE estimates agreed with that of independent studies and they argued that derived CPUE was a proxy for abundance. Certainly social media contributors will be more heterogeneous, but there is a potential for extracting many data points, particularly for commonly caught species and as social media adoption among the populace reaches saturation.

3.5.2 Trip Duration Estimates

Comparing the derived trip duration estimate with that of the 2016 MRIP survey (National Oceanic and Atmospheric Administration, 2016) of shore trips with striped bass prosecution indicates that social media may provide a credible estimate by which to scale effort (MRIP, 3.9 hours; Social Media Derived 3.9 hours). UK duration estimates from all other organised surveys (summary Table 3-3) exceeded 3.9 hours. Duration estimates from the Sea Angling 2012 on-site roving creel survey of shore fishers (abbr. SA2012; CEFAS, 2012; Armstrong *et al.*, 2013) for sea bass prosecuting trips was 5.1 hours, 95% CIs [4. 4, 5.7] (CEFAS, 2012) and all other Wales centric surveys (Goudge, Morris, & Sharp, 2009; Goudge et al., 2010; E A Richardson, 2006) had trip duration estimates exceeding 5 hours (all trips). MRIP is primarily an access point survey, hence it is not subject to the length-of-stay bias of roving creel surveys unlike SA2012. Striped bass were commonly subject to bag limits across the USA, which results in shorter trip durations when striped bass are the primary target species (Pollock, Hoenig, Jones, Robson, & Greene, 1997). Further evidence is required to establish that social media derived trip duration estimates can be more accurate than unadjusted access point surveys duration estimates. Other factors can contribute to the lower trip duration estimates observed from MRIP, most notably there were no bag limits in place for sea bass during the SA2012 survey. However, contrary to these results, it could be postulated that social media derived trip estimates would be inflated because self-selecting surveys under-represent trips without catches (e.g. Hartill, 2017) and no-catch trips tend to be shorter (this was the case in the SA2012 and 2016 MRIP datasets). The relative effect of competing inflationary and deflationary mechanisms is unclear and further work is therefore required.

Table 3-3. List of surveys which were discussed.

Title	Reference	Coverage	Description	Survey Instrument
Socioeconomic and ecological implications of an ecosystem approach to marine resource management for Wales, UK	Richardson (2009)	Wales, UK. 2003–2005	Doctorate thesis. Comprehensive geographically specific work for doctorate thesis, Bangor University. Includes very extensive effort and economic surveys with excellent coverage of for-hire boat sector and economic analysis of the recreational sector.	Online and face to face questionnaire survey instrument. Non-randomised as respondents were self-selecting.
North Wales Recreational Sea Angler Pilot Surveys: Winter 2007 and Summer Results July to October 2008	Goudge et al. (2009, 2010)	North Wales, UK. 2007–2008	Pilot survey commissioned by the Conservation Council for Wales (now Natural Resources Wales). Onsite survey aimed primarily at effort and catch assessment of shore angling in North Wales.	Questionnaire based non-randomised creel (on-site) survey
Sea Angling 2012	Armstrong et al. (2013)	England, UK. 2012	National Directed Survey, organised by CEFAS. First statistically rigorous sea angling survey in the UK. Multiple instruments were used in economic, effort and catch assessments. On-site survey did not cover Wales.	Questionnaire based stratified random on-site survey. Survey sites and survey times were randomly selected from a sampling frame.
Marine Recreational Information Program	NOAA (2016)	USA. 2016	A statistically rigorous survey program of catch and effort covering the USA which occurs annually and is overseen by the National Marine Fisheries Service (NOAA)	Dual frame stratified random mail survey to estimate effort.

3.5.3 Activity Patterns

The number of gears used by shore fishers per trip was statistically stationary between 2002 and 2013 and the dominance of rod-and-line gears was implicitly indicated by the failure of other gears to be identified in mined text. Gear use and modifications could be expected to increase in response to falling catches as observed in commercial fisheries (Marchal et al., 2007), certainly when equipment costs in the UK over the last two decades have seen a real-term reduction (Office for National Statistics, 2018). However, no change in gear numbers used per trip was detected, this is perhaps unsurprising as sea bass spawning stock biomass increased between 2002–2010 (ICES, 2017a). It could also indicate that the relatively low cost of equipment as a proportion of income meant that MRFs already deployed the maximum number of gears practical. This article did not specifically look for changes in capture methods, but social media data is certainly capable of providing information on fishing practices (e.g. Shiffman *et al.*, 2017) and has obvious parallels with recreational or subsistence hunting practices where technological creep could be an indicator of sector changes (e.g. Gill *et al.*, 2012). Social media text and images published to forums and equivalent entities (e.g. Facebook Groups) have been used to examine illegal practices (Di Minin et al., 2018; Eid & Handal, 2017; Shiffman et al., 2017; Siriwat, Nijman, Wildlife, & Sciences, 2018).

Seasonal cycles of sea bass prosecution agreed with the popular MRF literature. Surveys of the inshore commercial sector show a strikingly similar pattern of monthly effort (Pawson & Pickett, 1987; Pickett, 1990) and this pattern was repeated in the analysis of the SA2012 survey data (Armstrong et al., 2013a; CEFAS, 2012). During winter, sea bass older than ~2 years migrate from their summer feeding grounds to offshore areas (Holden & Williams, 1974; Pawson, Kelley, & Pickett, 1987; Pawson, Pickett, Leballeur, Brown, & Fritsch, 2007a) and hence larger fish are less frequently captured by inshore fishers. It may be expected that recreational fishing pressure on juveniles increases during winter as they remain inshore and available to the recreational fishery. However, this article showed that overall prosecution was significantly reduced. It is reasoned that effort is reduced because of a lower chance of landing trophy fish and fish over the minimum conservation reference size which can be harvested for consumption. Catchability using baits and bait imitations almost certainly decreases as the sea temperature falls, with an associated drop in the metabolic rate of sea bass and a reduction in their food consumption (Claireaux & Lagardère, 1999; Person-Le Ruyet, Mahé, Le Bayon, & Le Delliou, 2004). It is likely that this reduction in catches contributes to switches to colder water species such as the Atlantic cod as indicated by the analysis of the Sea Angling 2012 survey dataset (CEFAS, 2012).

Given that this temporal pattern was detected in the mined data it indicates that social media is capable of providing relative and qualitative information on temporal patterns of MRF effort. This prosecution effort also correlated with seasonal changes in the abundance of adult sea bass inshore however, the factors which drive seasonal effort are complex and further work would be required to disentangle the relative contribution that social factors (e.g. increased tourism in summer) and sea bass availability make to changes in effort.

Georeferenced data mined from social media enabled reporting of seasonal changes and spatial patterns of sea bass prosecution beyond those previously available. This extends the work of Martin et al. (2014) which used keyword frequencies counts found in forum texts as a proxy for recreational fisher effort on freshwater lakes. No previous survey of MRFs had complete spatial coverage of Wales (to September 2018) and the only data published on the spatial and temporal distribution of MRFs who prosecute sea bass from the shore was derived from expert knowledge (Pawson et al., 1987). Previously, only Goudge *et al.* (2009, 2010) carried out on-site surveys which recorded catch by species (including sea bass) but sea bass catches accounted for only 0.001% of total recorded catches. Unfortunately this lack of data for sea bass prosecution from the shore makes validation of the spatial results problematic. Geographic areas of increased sea bass prosecution were broadly coincident with primary road access and proximity to population centres. North Wales had the greatest change in prosecution between summer and winter and it is notable that North Wales also has the greatest increase in tourism visits (39% total visit share) during the summer months (TNS Global, 2014a).

3.5.4 Limitations and Bias of Social Media Activity Data

The primary barrier to the acceptance of results derived from social media data and other participant knowledge sources is the unquantifiability of bias. Soundly designed surveys which seek to quantify wildlife harvest use random multi-frame sampling to provide unbiased estimators of key population metrics (e.g. harvest and participation rates) across strata chosen to minimise sample variance. Multiple frames are used to correct estimators derived from incomplete frames. The statistical moments calculated from social media data have no true sampling frame hence population estimators cannot be corrected by randomised sampling from a complementary non-overlapping frame. Hence it is unclear how any bias correction can be applied using accepted statistical techniques used in surveys (Kish, 1995). It might be possible to determine bias direction and magnitude empirically where statistically sound surveys have a similar scope, but generalizability will be limited by that scope. It will be difficult to identify 'real world' respondents who are outside the notional sampling frame (e.g. forum users whose probability of reporting a trip is non-zero) from which the social media trip reports were sampled. Hence it would be incorrect to generalize the results to a wider population (e.g. using a harvest per unit time estimator to derive total population harvest).

Careful consideration must be given to which information can usefully be derived from social media and how that information is communicated to stakeholders and there are many challenges in using social media in conservation science (Di Minin, Tenkanen, & Toivonen, 2015). Social media data is subject to errors of representation (Groves et al., 2009), as are all self-selecting survey instruments. Intentional deception also occurs in social media (Tsikerdekis & Zeadally, 2014). Prestige biases are particularly relevant in the present study (Campbell et al., 2001) and to trophy hunting, which will manifest as (i) exaggeration of size (ii) over reporting of trophy specimens (iii) under reporting of mundane trips where only common species, non-trophy specimens or no captures were made. Type (i) prestige bias could be limited where photographic evidence is provided and the mining process could include scraping associated photographs. Other verification methods include the review of the post history of individuals to flag a record of erroneous reports and statistical methods to detect outliers. Recent gamification of social media provides an opportunity to use user feedback to grade content trustworthiness (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008; Bian, Liu, Zhou, Agichtein, & Zha, 2009).

Measurement and species identification errors would be difficult to identify unless an image accompanies the content. This approach was used by Shiffman et al. (2017) to verify shark species captured by recreational anglers. The necessity to process unstructured text when handling social media data will increase error rates compared to questionnaire led surveys. Despite these problems, social media mined data reveals the preferences of participants without biases which arise when soliciting a response directly. It is reasonable to assume that recall bias will be reduced in comparison to some survey instruments as posts will generally be made soon after the trip occurred. Other biases which are associated with elicited responses (e.g. deference and order effects) should also be reduced (Fowler Jr & Mangione, 1990; Newing, Eagle, Puri, & Watson, 2010). The problem which arises is quantifying the size and direction of these biasing effects to make reasonable corrections, it is conjectured that such a correction is not possible in practice.

Additional statistical problems exist in making population inferences from social media derived sample estimates beyond those exhibited by self-selecting survey instruments (Heckman, 1990; Lavrakas, 2008). In the present study, Region NE2 had reported activity twice that of any other region, but this may not be a real-world difference. Social media data are non-independent in space and time and a small proportion of users within the online community may tend to provide repeated contributions (Lerman, 2007; Nielsen, 2017; van Mierlo, 2014). Clearly the locations frequented by participants in their recreational activity will not be randomly chosen. Social media posts are likely to influence others in the social network (Bond et al., 2012; Centola, 2010) and will increase contributions and may stimulate recreational activity in other users. Different social media communities could be treated as units of replication however, users can contribute pseudonymously to more than one site and network effects (Katz & Shapiro, 1994) drive social media users to use fewer sites. The bulk of recorded trips mined in the present study were derived from a single discussion forum and other forums were data sparse.

SNSs typically allow social media users to volunteer their demographic data. This information could be compared with randomised surveys to determine how representative sampled social media users were of the population. This presents two problems. Firstly, if user anonymity were broken then retaining such demographic data could qualify the research to involve human subjects, which would then require informed consent which is impractical or impossible (Monkman, Kaiser, et al., 2018). Secondly, demographic matching (e.g. on age or avidity) between social media derived data and other surveys will clearly not produce population estimators to which statistical certainty can be reliably assigned. The notional sampling frame of social media samples is unknowable and using demographic matching to derive population estimates would fall foul of an ecological fallacy or a fallacy of composition.

3.6 CONCLUSION

The present study demonstrates that text and data mining of publicly accessible social media content can reveal patterns of participation in wildlife recreation which were only previously inferred from expert and participant knowledge. The volume of mined data can exceed that obtained from general directed surveys where prosecution activity is comparatively rare in the broader population, or where encounters with participants are unlikely as part of a general survey. Encounters could be rare because of the remoteness of where the activity occurs, because the activity occurs during unsocial hours (Armstrong et al., 2013a; National Research Council, 2006) or because of participant secrecy (Maurstad, 2002). It is likely that such characteristics are shared with other wildlife recreation activities (Olsen & Thuen, 2013).

Conventional assessments of recreational activity describe participant behaviour across a single time span. Perhaps the most promising use of online social media is in near real-time monitoring of shifts in the behaviour of participants in wildlife recreation particularly in new and developing sectors and under the increasing adoption of social media through low cost access to the internet from mobile phones in developing nations (G. Zhang, 2017). Opportunities may exist to use social media to investigate shifts in population structure from trophy records (Elizabeth A Richardson et al., 2006) and to contribute to presence-absence datasets where multiple data sources are beneficial (Lepczyk, 2005) and help eliminate false absences. Social media monitoring has been used with success to track disaster events (review Alexander, 2014) and in epidemiology (review Eysenbach, 2011), although there have been high profile failures (Lazer, Kennedy, King, & Vespignani, 2014). There is the potential to explore socioeconomic questions, such as how participant demographics change and how participants respond to management and policy changes using culturomics and sentiment analysis (Ladle et al., 2016; Palomino, Taylor, Göker, Isaacs, & Warber, 2016), although ethical issues and the nature of social media itself could make this problematic (Monkman, Kaiser, et al., 2018).

Chapter 4

Heterogeneous Public and Local Knowledge Provides a Qualitative Indicator of Coastal Use by Recreational Fishers

Chapter 4 has been accepted and is in-press with the journal *Journal of Environmental Management*

https://www.journals.elsevier.com/journal-of-environmental-management

KH and MK critically reviewed and revised the article. All other work was that of GGM.

Highlights

- Heterogeneous local and public knowledge can provide high resolution maps of the spatial distribution of coastal use by marine recreational fishers across national scales for management and marine spatial planning.
- Site density as a simple proxy of activity performed at least as well as more complex indicators.
- Social media can provide higher volumes of local knowledge data than other participant and local knowledge sources.
- Applying a linear focal mean to activity maps improves agreement with independent survey results.
- While local knowledge can provide qualitative activity maps, the mined data has inherent unknowable biases.

4 Heterogeneous Public and Local Knowledge Provides a Qualitative Indicator of Coastal Use by Marine Recreational Fishers

4.1 Abstract

Marine recreational fishing (MRF) benefits individuals and economies, but can also impact fish stocks and associated ecosystems. Fish are an important resource providing direct economic benefit through commercial and recreational exploitation, and more esoteric ecosystem services. It is important to consider recreational fishing in marine spatial planning, but spatial information on coastal utilization for MRF is frequently lacking. Public sources of local knowledge were reviewed and the frequency of unique references to sites extracted. Sites were georeferenced using a gazetteer compiled from the Ordnance Survey and United Kingdom Hydrographic Office named sea features gazetteer and local knowledge sources. Recreational fishing site densities were calculated across 2,700 km of coastline and this proxy indicator of coastal utilization validated against two independent surveys using permutative Monte Carlo sampling to control for sparse and non-independent data. Site density had fair agreement with independent surveys, but standardization by shore length reduced this agreement. Applying a 3 by 3 box filter convolution to the spatial layers improved the agreement between local knowledge derived predictions of activity and those of directed surveys, and permutation testing showed that agreement did not arise as a result of the convolution itself. High and low activity areas were more accurately predicted than areas of intermediate activity. Site density derived from heterogeneous participant and local knowledge can produce qualitative predictions of where recreational fishers fish, and applying a convolution can improve the predictive power of data so derived. However, this approach will be subject to unquantifiable bias and may fail to identify areas highly valued by marine recreational fishers. Thus it should be used in conjunction with other information in decision making and may be best suited to inform the early stage sampling design of on-site surveys or to complement other data sets in mapping areas of importance to recreational fishers.

4.2 INTRODUCTION

Coastal and marine spatial planning (MSP) frameworks have become an integral tool in the governance of marine and coastal resources in the European Union, America and many other nations (e.g. European Commission, 2014; MaPP, 2016; The White House, 2010; Vince, 2014). In Europe, the European Parliament has adopted Directive 2014/89/EU to create a common framework for maritime spatial planning (European Commission, 2014) and the USA has adopted the National Policy for the Stewardship of the Ocean, Our Coasts, and the Great Lakes of USA (The White House, 2010). MSP aspires to achieve the equitable allocation of marine and coastal resources where stakeholder activities are potentially in conflict. The aim of using the MSP framework is to ensure that benefit maximisation occurs now and in the future and is considered a vital component of ecosystem-based management (Douvere, 2008; Environmental Law Institute, 2009).

Fisheries are an important marine resource used by humans for both food production and recreation. For this reason, MSP should evaluate the interaction among those marine and coastal stakeholder activities that might impact marine fisheries. Historically, marine recreational fishing (MRF) was considered to have negligible impact on fisheries hence recreational harvests have been omitted from stock assessments of commercially important species. This orthodoxy has changed and contemporary research has demonstrated the potentially large numbers of fish caught by recreational fishers (Coleman et al., 2004; Hyder et al., 2018; Post et al., 2002; Z. Radford et al., 2018) and the possible ecosystem and environmental effects associated with MRF (Hyder et al., 2017; Lewin et al., 2006; O'Toole, Hanson, & Cooke, 2009). For these reasons, there is increasing interest in trying to include MRF data in stock assessments (e.g. Eero et al., 2014; Hyder et al., 2017; ICES, 2017a).

It has been suggested that unaccounted recreational harvest can impede stock recovery in managed fisheries (Maggs, Mann, Potts, & Dunlop, 2016; Sherwood & Grabowski, 2016). Conversely, research has identified the benefits of MRF to economies at national and local levels (A. Brown et al., 2013; Donnelley, Radford, Riddington, & Gibson, 2009; Gartner et al., 2002; Herfaut et al., 2012; Roberts et al., 2017) and in health and wellbeing (A. Brown et al., 2012; Gartner et al., 2002; Griffiths et al., 2016; Lawrence & Spurgeon, 2007). Balancing the interests of marine recreational fishers (MRFs) with ecological considerations and other marine stakeholders is therefore an important aspect of MSP. The potential importance of the sector has been recognized, with recreationally important stocks protected from commercial exploitation to assure the quality of recreational fishing (e.g. Irish Parliament, 2006; Isle of Man Government, 2016; Maine Department of Marine Resources, 2016)

Despite the importance of MRF, many countries do not undertake regular assessments of recreational fishing activity. For example, only four European member states have recreational mortality estimates for European sea bass (*Dicentrarchus labrax*, henceforth sea bass) (Hyder et al., 2018; ICES, 2017d). This is despite evidence that recreational sea bass catches can be significant (Armstrong et al., 2013a; Herfaut et al., 2010; Rocklin, Levrel, Drogou, Herfaut, & Veron, 2014; van der Hammen & de Graaf, 2015), concerns over stock health, and lack of data for stock assessment (ICES, 2017a, 2017b). Moreover surveys to estimate MRF effort and catch at a national level tend not to provide the level of sampling needed to produce detailed information on the spatial distribution of activity (Armstrong et al., 2013a).

Directed surveys to assess catch frequently use on-site access point or roving creel methodologies to assess catch (Guthrie, 1991; National Research Council, 2006; Pollock et al., 1997). Random sampling is frequently achieved by including location (site) as a randomly sampled component, yet the sampling frame of sites will not represent 100% spatial coverage of the entire coastline or all access points. Expert knowledge and pre-survey scoping can be used to create a sampling frame (e.g. Armstrong et al., 2013a) where activity is known to occur and this may include proportionate sampling based on expected site popularity. The development of site sampling frames is improved by considering all available data sources prior to the finalisation of the sampling regime.

In the absence of directed surveys, several methods have been used to assess MRF activity in data poor fisheries. Self-selecting and non-randomised surveys are commonly employed (e.g. Aron et al., 2014; Drew Associates Ltd., 2004; Goudge et al., 2010, 2009; McMinn, 2013). However, spatial coverage is limited by the spatial distribution and number of volunteers, or by site selection criteria. Expert and local knowledge are an important information source (Hind, 2014, 2015; Johannes et al., 2000) and can be the best available information in emerging and artisanal fisheries (e.g. Deepananda, Amarasinghe, Jayasinghe-Mudalige, & Berkes, 2016; Stange, 2016). The past decade has seen an increase in engagement between recreational fishers and researchers as co-management is increasingly recognised as being important for long-term and effective management (review Linke and Bruckmeier, 2015). Smartphones and social media provide a means of both delivering and promoting software which allow recreational fishers to record catch and other fisheries observations which can be used by scientists involved in fisheries research (review Venturelli, Hyder, & Skov, 2017) and co-management approaches.

MRF records can be used to derive trends in stock status and MRF activity levels (Barbini et al., 2015; Bennett et al., 1994; Gartside et al., 1999; Elizabeth A Richardson et al., 2006). However, accessible data repositories tend to be held by MRF clubs or hobbyist magazines that are unrepresentative of overall activity when considered in isolation. It is apparent that heterogeneous data sources exist from which fisheries data can be derived, but spatial coverage will be limited according to the spatial distribution of clubs and other contributory sources. By combining multiple sources, it is expected that detailed maps of the relative levels of spatial activity could be produced which can be used to inform management and the marine spatial planning process where data is lacking.

Here I use a case-study of a data poor recreational fishery (Wales, United Kingdom) to show how heterogeneous knowledge sources can be used to produce spatial indicators of shore use by MRFs for the purposes of marine spatial planning (see UK Marine Policy Statement), or in other information gathering exercises. I compare several proxy measures for coastal utilisation by MRFs and validate their performance against two independent directed surveys. The best performing activity proxy is further analysed using novel permutative Monte Carlo sampling to determine the suitability of the proxy as an indicator of coastal utilisation.

4.3 METHODS

4.3.1 SCOPE

The scope of the present study was recreational fishing on the coastline of Wales, UK. This article's definition of recreational fishing accords with that of Pawson *et al*. (2008). Only fishing with rod-and-line (angling) was considered as this method dominates activity in England (Armstrong et al., 2013a) and there were no instances of other fishing methods recorded in the literature. Much of the 2,740 km of Welsh coastline was accessible to MRFs. The term public and local knowledge refers to all publicly available sources in which spatial data on coastal use by MRFs was published. Local knowledge means locale specific information published by fishers with respect to MRF activity.

The Welsh Government is responsible for the management of its waters and has obligations to report harvest estimates of some recreational catches under the European Union's multi-annual programme for data collection (EU Decision 2016/1251). Obligations also exist concerning equitable and optimal use of marine resources and good management of the marine environment under the Marine and Coastal Access Act 2009 (UK Parliament, 2009) and the Welsh Government are currently committed to producing the Welsh National Marine Plan (Welsh Government, 2017) under EU directive 2014/89/EU to establish a framework for maritime spatial planning.

4.3.2 SOURCE IDENTIFICATION AND DATA ACQUISITION

Sources recording MRF sites were classified as sea angling literature, social media used by MRFs, government related assessments and academic research. The www.google.co.uk search engine was used in October 2014 to identify angling literature, social media, and government commissioned assessments which may detail sites of MRF activity. The Google search terms were *(Wales OR Welsh) AND (angling OR fishing) AND (sea OR marine).* The scientific literature was searched using Google Scholar (scholar.google.co.uk), Web of Science (apps.webofknowledge.com) and ProQuest (search.proquest.com) using logically equivalent search terms. All relevant sources were recorded (Appendix E). Sources were reviewed for the presence of sites used by MRFs in Wales. Some data sources only had partial coverage of Wales (e.g. some were dedicated to fishing in Pembrokeshire in South Wales). It was expected that the number of sources covering a spatial area (*coverage count*) would need to be accounted for in activity estimates, hence coverage extents were created during geoprocessing so the number of contributing sources at any point were known. All data were anonymised and stored in an encrypted Microsoft SQL Server database (Microsoft, 2008).

4.3.3 Georeferencing and geoprocessing

To determine the geographical coordinates of MRF sites found in sources, it was necessary to compile a gazetteer of place names with their latitude and longitude. The gazetteer was compiled by merging the following: (i) all settlement names within 3 km of the Wales coast in the Ordnance Survey (OS) gazetteer (Ordnance Survey, 2015); (ii) all UK Hydrographic Office (UKHO) named sea features (United Kingdom Hydrographic Office, 2013) in Welsh national waters; and (iii) colloquial place names in local knowledge sources (e.g. popular angling literature and social media). The finalized gazetteer consisted of 6,610 named coastal and marine locations with 5,536 (84%) from the OS and UKHO, and 1,074 (16%) from local knowledge sources.

Where necessary, all conversions between British National Grid and WGS84 used the OSGB 1936 WGS 1984 Petroleum transformation. Some geoprocessing tasks were performed with ET GeoWizards (E T Spatial Technologies, 2014) and the Geospatial Modelling Environment 0.7.2 RC2 (Beyer, 2015). All spatial analysis used a regular 1 km^2 vector grid (henceforth *cells*) which were aligned with the FishMap Môn cells (FishMap Môn is described later). Only cells that intersected the mean high water line were retained and the centroid of gazetteer locations were snapped to the mean high water line to ensure that matched sites were coincident with cells for all processing.

Where any metric was expressed as a unit of shore length, the topological complexity of the coastline was smoothed in ArcGIS 10 by applying a polynomial approximation with exponential kernel (PAEK) smoothing with 100 metre tolerance to the mean high water line. The output was reviewed for locations known to the authors, to validate the removal of 'meso level' shore features < 20 m features, while preserving features > 20 m. In estuarine areas the high water line can extend many miles in land, all estuarine areas were truncated where the estuary width was under 100m. Henceforth all references to shore length refer to smoothed shore length.

4.3.4 Calculation of Use Proxies

Three different proxies for coastal use were calculated for each cell as follows.

(i) *Site density*, d, the number of fishing sites within a 1 km^2 cell, see Figure 4-1 for a detailed explanation.

(ii) *Standardized site density*, $\acute{d}$, $\acute{d} = d/l$ where l is the smoothed shore length of mean high water springs, (e.g. 2 km in Figure 4-1).

(iii) *Coverage score*, s, which adjusts for the coverage count falling within a 1 km^2 cell given by $s = 0.5^{k-x}/0.5^{x} \cdot l$, where k is the coverage count, x is site density for the cell and l is smoothed shore length. Outliers were set a ceiling value of mean +2 standard deviations (SD, s = 16.82), this was 0.01% of non-zero 1 km^2 cells.

4.3.5 VALIDATION

Validation required independent sources of MRF activity data covering the same spatial and approximate temporal scope. These sources were compared against the best performing proxy identified from analysis in the present study. Sources should have transparent and systematic methodologies and publicly available results. Two such studies met these criteria, the FishMap Môn (FMM) project (Aron et al., 2014), and the Wales Activity Mapping (WAM) project (Chambers, Haines, & Pitts, 2013; D. Jones, 2017). The geographical areas of FMM and WAM are given in Figure 4-2 and methodological summaries of the projects follow. The general term *validation* applies to the FMM and WAM datasets, data derived from local knowledge will be referred to as *test.*

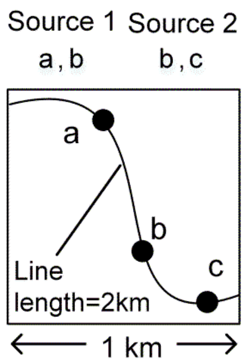

Figure 4-1. Calculating site density *d*.

Consider two sources 1 and 2 (e.g. the venues section of an online fishing forum and a book on sea fishing in Wales). Source 1 names sites *a* and *b* as fishing locations, and source 2 names locations b and c. The number of occurrences of a site name within a source is not considered. The site density in this example is 4 sites km^{-2}. The standardized site density $\acute{d}$, which accounts for the length of coastline within the cell is 4/2 km, i.e. 2 sites km^{-1}.

4.3.5.1 FISHMAP MÔN

In 2013, FMM piloted methods in the collection of fishing activity data covering Anglesey and the surrounding coastline. FMM shore MRF activity (henceforth FMM intensity, I_{FMM}) was mapped using a creel survey. Natural Resources Wales identified 43 sites split among seven regions from which survey sites were selected without randomisation for surveyor visits. MRFs were asked to mark a map with their fishing locations. FishMap Môn provided pre-aggregated data for 1 km^2 cells in units quoted as angler visits hectare^{-1} week^{-1}.

4.3.5.2 WALES ACTIVITY MAPPING

In 2013, WAM mapped marine based recreational activity and associated economic value across Pembrokeshire using face-to-face and telephone interviews with regional experts (“wardens, rangers, outdoor centre instructors, recreation managers, beach managers and harbour masters”) involved in marine recreation (D. Jones, 2014). The selection criteria for participants was not specified however, it is reasonable to assume that web searches, Welsh Government employee lists, snowballing and personal knowledge were used to collate a survey frame. Surveyors asked interviewees to mark areas of activity on a map. An indication of the frequency of use was recorded by asking the interviewee to classify the number of site visitors and the days in a month subject to ‘moderate activity’ on two ordinal scales for 5 separate annual periods. The method of Chambers *et al*. (2013) was used to convert these data to an annual estimate of total fishing days a year (henceforth WAM intensity) occurring within the polygonised areas (WAM polygon).

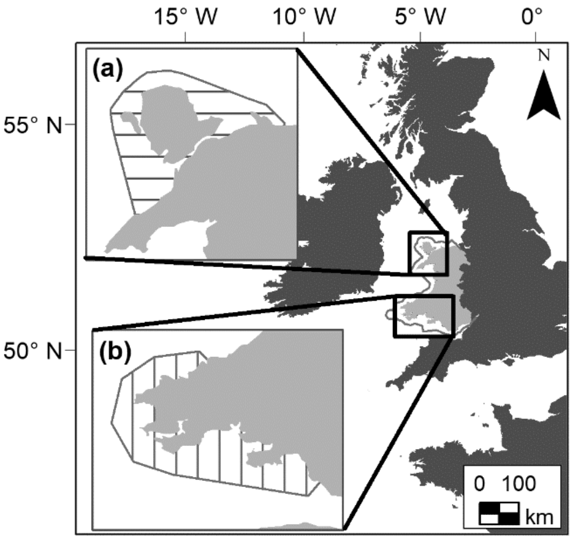

Figure 4-2. Hatched areas of insets are marine recreational fisher survey extents for (a) FishMap Môn and (b) Wales Activity Mapping.

The feint grey area in the main map is Wales, UK and the outline around Wales is the 6 nautical mile limit.

Cells can contain many WAM polygons and vice versa. To calculate the activity within a cell, the level of activity assigned to an individual polygon was weighted according to the area of the polygon falling within the cell. Figure 4-3 illustrates a cell with 2 intersecting WAM polygons and formally, the WAM intensity within a cell (henceforth I_{WAM}, with units of trips year^{-1} km^{-1}) was calculated according to $I_{WAM} = \frac{1}{l} \cdot \sum_1^n \frac{I_i \cdot a_i}{A_i}$, where i is the i^{th} intersecting polygon, l = shore length, I = polygon intensity, a = polygon area within the cell, and A = polygon area (e.g. $a_1 + b_1$ in Figure 4-3).

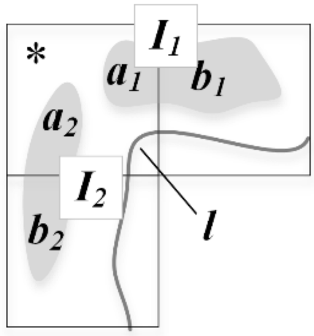

Figure 4-3. Cell* intersects two polygons which are assigned intensities of I_1 & I_2, where I_1 and I_2 are the number of fisher days per year occurring in the whole polygon area. Each polygon contributes to the fishing intensity of cell*, by weighting the intensity of the polygon by the polygon areas a_1 and a_2 within cell*. l = shore length within cell*.

4.3.5.3 *General statistical methods*

Correlations between the 3 use proxies (d, $\acute{d}$, s), and FMM and WAM intensities were tested with Kendall's rank correlation tau-b (Kendall's tau-b) as data were nonparametric and had tied ranks (Howell 1997, p. 293). The inter-rater agreement between site density, and FMM and WAM intensities was calculated using an equally weighted Cohen's kappa (Cohen, 1968) after all data had been converted to quartiles. The Python package SciPy ver. 0.18.1 (E. Jones, Oliphant, Peterson, & others, 2017) was used to calculate Kendall's tau-b test statistic. P-value estimation for non-permutative tests were calculated in R (stats:cor[3]) because SciPy ver. 0.18.1 misreported the P-value. Cohen's kappa testing used the Python package StatsModels (Seabold & Perktold, 2010). An α of 0.05 was used for all significance tests.

4.3.5.4 *Additional processing and permutation tests*

Several problems may confound the reliability and interpretation of spatial correlation tests as follows.

- (i) Data are unlikely to be independent.
- (ii) The interpretation of standard tests of correlation and inter-rater agreement is unclear in the presence of many spatially matched zeros (paired zeros), and other ranks, or when matched zeros are excluded from tests. Figure 4-4c illustrates the issue of spatially matched zeros.
- (iii) Sparse data with paired zeros will greatly inflate tests of correlation and inter-rater agreement.
- (iv) In comparing exactly coincident cells between two layers, no allowance is made for small spatial differences which arise between the two tested layers as a result of the original data collation process and processing.

[3] Validated as Kendall's tau-b by comparing outputs with SPSS v20 (IBM Corp, 2011) and Real Statistics (Zaiontz, 2017).

(v) Applying a box filter (outlined below) to two spatially coincident data sets will tend to increase the correlation.

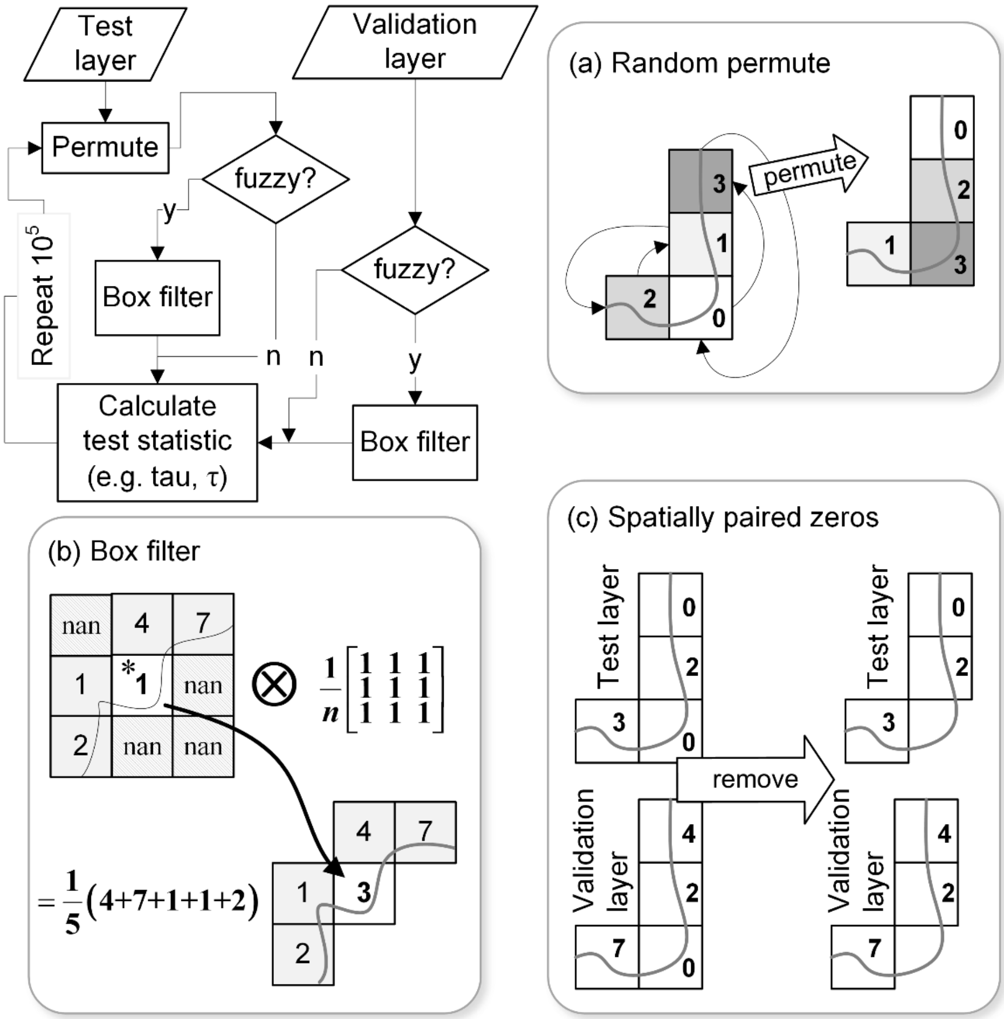

Figure 4-4. Outline of Monte Carlo permutative testing process, applied to Kendall's rank correlation tau-b and Cohen's kappa. The test layer contains cells assigned a site density (sites km^{-2}) as calculated in the present study. The validation layer will be derived from the FishMap Môn (Aron et al., 2014) or the Wales Activity Mapping (Chambers et al., 2013) projects. The test layer was randomly permuted while maintaining the spatial relationships of cells (see [a]). Statistical tests compared values shared by precisely coincident cells (crisp) and after applying a 3x3 box filter (fuzzy, [b]) to each cell with a value (nan = not a number). Tests were also carried out after excluding spatially paired zeros, as illustrated in (c). The line within the cells represents the shoreline.

To overcome issues (i), (ii) and (v) Monte Carlo permutation tests were used to generate estimates of p (Odiase & Ogbonmwan, 2007). The interpretation of permutation tests is conceptually easier, it is the probability of achieving a higher correlation from within the sample than that observed between the two original layers. The null hypothesis remains the same, i.e. that there is no correlation between the two variables however, using a permutative approach allows the estimation of p with no prior assumptions of the probability density function of the observations. The first step is to calculate the test statistic for the original layers. One layer is then permuted and the test statistic recalculated. The probability p of getting a stronger positive correlation is estimated according to $p = \frac{1}{n} \cdot \sum_{1}^{n} \begin{cases} 1, f_i > F \\ 0, f_i \leq F \end{cases}$, where f_i represents the i^{th} test statistic (n = 100,000) and F is the test statistic calculated from the original validation and test data. Issue (iii) was addressed by rerunning the tests after excluding paired zeros (definition, see Figure 4-4c). Issue (iv) was addressed by applying a box filter (aliases, focal mean or linear convolution) to each cell in both layers as follows, $\frac{1}{9} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, noting that cells within the bounds of the kernel without data are excluded and the fractional term adjusted accordingly (see Figure 4-4b). The permutation process flow is outlined in Figure 4-4. Henceforth, results derived from exact spatial matches will be known as *crisp*, those based on box filtered cells are *fuzzy*. Spatial permutations and related calculations were written by the authors and made extensive use of the Python package NumPy (van der Walt, Colbert, & Varoquaux, 2011).

The cumulative proportion p_{ij} of quartiles or tertiles (for brevity, simply quartiles) was calculated according to $p_{ij} = \frac{1}{x_i} \cdot \sum_{0}^{j} y_j$. Where i is the quartile of the validation survey (i.e. FMM or WAM, e.g. $i = \{0, 1, 2, 3, 4\}$ for quartiles) and j is the quartile separation, e.g. $j = \{0, 1, 2, 3, 4\}$. Let x_i be the count of cells in the directed survey (FMM or WAM) of quartile i and y_j be the count of cells at a separation of i. *Separation* is defined as the number of quartiles between i and j as illustrated in Figure 4-5.

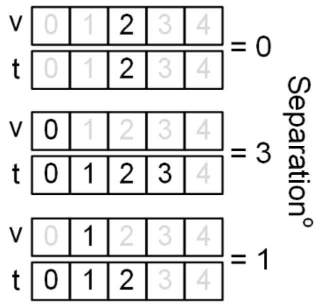

Figure 4-5. Quartile separation calculation examples.

v=validation, t=test. In calculating the cumulative quartile proportion all quartiles frequencies ≤ separation° are included. Where a separation of 1 or 2 would result in inclusion of two additional frequencies, a random one of the two was first selected.

4.4 RESULTS

4.4.1 SOURCES

Searches identified 57 information sources relating to MRF activity (Appendix E), 25 (44%) of these had high resolution spatial data. Of these 25, user generated content published on social media was the most prevalent, with 14 (56%) distinct sources. Appendix F summarises the use of relevant data sources and the reasons for exclusion of unused sources. Only 2 (5.1%) sources recorded non-rod-and-line activity (e.g. netting or spear fishing) and these sources had very low volumes of relevant content. Shore-based MRF activity was better represented in the sources than private boat activity (shore platform, n = 21 [84%]; private boat platform, n = 2 [8%]; shore & private boat, n = 2 [8%]) with social media local knowledge sources almost devoid of detailed spatial data on private boat MRF activity for Wales.

The locations of site data were published in popular MRF books and magazines, online forums, blogs and web sites. Other online sources included public Google Maps layers and downloadable GPS files. The temporal and spatial coverage of published local knowledge sources was highly variable, with literature derived records available prior to the 1980s and social media sources only available after 2001. Of the 25 sources identified, 9 (36%) covered all of Wales, 6 (24%) of which were extracted from web content. The remainder were dedicated to regional areas, such as North Wales, Anglesey and the Pembrokeshire coastline.

4.4.2 SPATIAL DISTRIBUTION AND VALIDATION

Across sources, 2,700 occurrences of locations were matched with sites contained in the gazetteer. Of the 2,700 locations, 1,223 (45%) were from traditional published literature, 878 (33%) were from social media and 599 (22%) were derived from directed surveys. Of the 1,558 1 km^2 cells intersecting the mean high water mark (MHW) across Wales, 974 (60%) had no recorded occurrence of an MRF site in any source. With zeros excluded, site numbers per kilometre of MHW length ranged between 0.02 locations km^{-1} at the eastern limits of the Bristol Channel and 10 locations km^{-1} on the western shore of Carmarthen Bay (median = 3.00, interquartile range: 0.74 – 5.00, n = 1244). Site densities d ranged between 0 and 10 sites km^{-2} (zeros excluded, median = 2, interquartile range: 1 – 4, n = 621). Across Wales, densities tended to be higher in South Wales (Figure 4-8) and Mid Wales tended to have the least activity, with the exception of some localised high activity areas concentrated around urbanised regions (e.g. Aberystwyth).

Coverage score (s) and shore length standardized site density ($\acute{d}$) were poor predictors of crisp sans zero ordinalized activity derived from FishMap Môn (Kendall's tau-b, *s vs.* FMM, $\tau = 0.03$, p = *0.68;* $\acute{d}$ *vs.* FMM, $\tau = -0.03$, p = 0.64). For the Wales Activity Mapping (WAM) survey, s performed poorly as an activity proxy for south west Wales ($\tau = 0.03$, p = 0.57) however, $\acute{d}$ did correlate with WAM activity ($\tau = 0.57$, p = 0.03). All correlation tests which included spatially paired zeros (definition Figure 4-4) were significant ($p < 0.00001$). Site density d had the strongest agreement with FMM ($\tau = 0.12$, p = *0.01*) but performed poorly as a predictor of activity for WAM ($\tau = -0.05$, p = 0.13) however, applying Occam's razor, all proceeding results are based on the simpler activity proxy d.

Viewing the subfigures (a) to (b) and (i) with (ii) of Figure 4-6 the predominance of 1 km^2 cells in which no activity was detected is apparent. The validation layers had fewer cells in which no activity was found (WAM, 52% zeros; FMM, 45% zeros) when compared to the test layers derived from public and local knowledge (WAM, 72% zeros; FMM, 60% zeros). This predominance of paired zero cells increased the agreement between intensity and d between the validation and test layers. Test and validation layers did not concord on repeating the correlation and inter-rater agreement analysis with paired zeros removed ($p > 0.05$), except under permutation testing of the FMM correlation ($p < 0.05$) and the inter-rater agreement for WAM ($p < 0.02$). Comprehensive statistical results of inter-rate agreement and correlation tests appear in Table 4-1.

Table 4-1. Kendall's tau-b correlation and Cohen's Kappa inter-rater agreement (IRA) between site density derived from local knowledge and the FishMap Môn (FMM) and Wales Activity Mapping (WAM) directed surveys. Kappa was calculated following conversion of activity measures to quartiles. *Standard*, results from conventional tests. *Permuted*, *p* estimate from 100,000 random permutations of the original cells, n_{ext} is the number of permutations with a test statistic more extreme than the standard. *Fuzzy* results report agreement after applying a 3x3 equally weighted box filter. *Zero*, all spatially coincident cells were included; *No Zeros*, coincidence cells with no recorded activity were excluded. Key for p: * $0.05 \geq p > 0.01$, ** $0.01 \geq p > 0.001$, *** $p < 0.00001$. Key for IRA: † fair, $0.2 < K \leq 0.4$ (Landis & Koch, 1977).

Kendall's tau-b correlation

		Crisp		**Fuzzy**	
Source	**Test**	**Zeros**	**No Zeros**	**Zeros**	**No Zeros**
FMM	Standard	$\tau = 0.335$ $p < 0.00001$***	$\tau = 0.120$ $p = 0.099$	$\tau = 0.381$ $p < 0.00001$***	$\tau = 0.329$ $p < 0.00001$***
	Permuted n = 100,000	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 4949$ $p = 0.049$*	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 0$ $p < 0.00001$***
WAM	Standard	$\tau = 0.272$ $p < 0.00001$***	$\tau = -0.045$ $p = 0.125$	$\tau = 0.250$ $p < 0.00001$***	$\tau = 0.147$ $p < 0.00001$***
	Permuted n = 100,000	$n_{ext} = 0$ $p = 0.00001$***	$n_{ext} = 87472$ $p = 0.875$	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 280$ $p = 0.003$**

Cohen's Kappa Inter Quartile/Tertile Ranges

FMM	Standard	$K = 0.261$†	$K = 0.040$	$K = 0.354$†	$K = 0.238$†
	Permuted n = 100,000	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 29{,}806$ $p = 0.298$	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 105$ $p = 0.001$**
WAM	Standard	$K = 0.231$†	$K = 0.155$	$K = 0.278$†	$K = 0.273$†
	Permuted n = 100,000	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 1816$ $p = 0.018$*	$n_{ext} = 0$ $p < 0.00001$***	$n_{ext} = 0$ $p < 0.00001$***

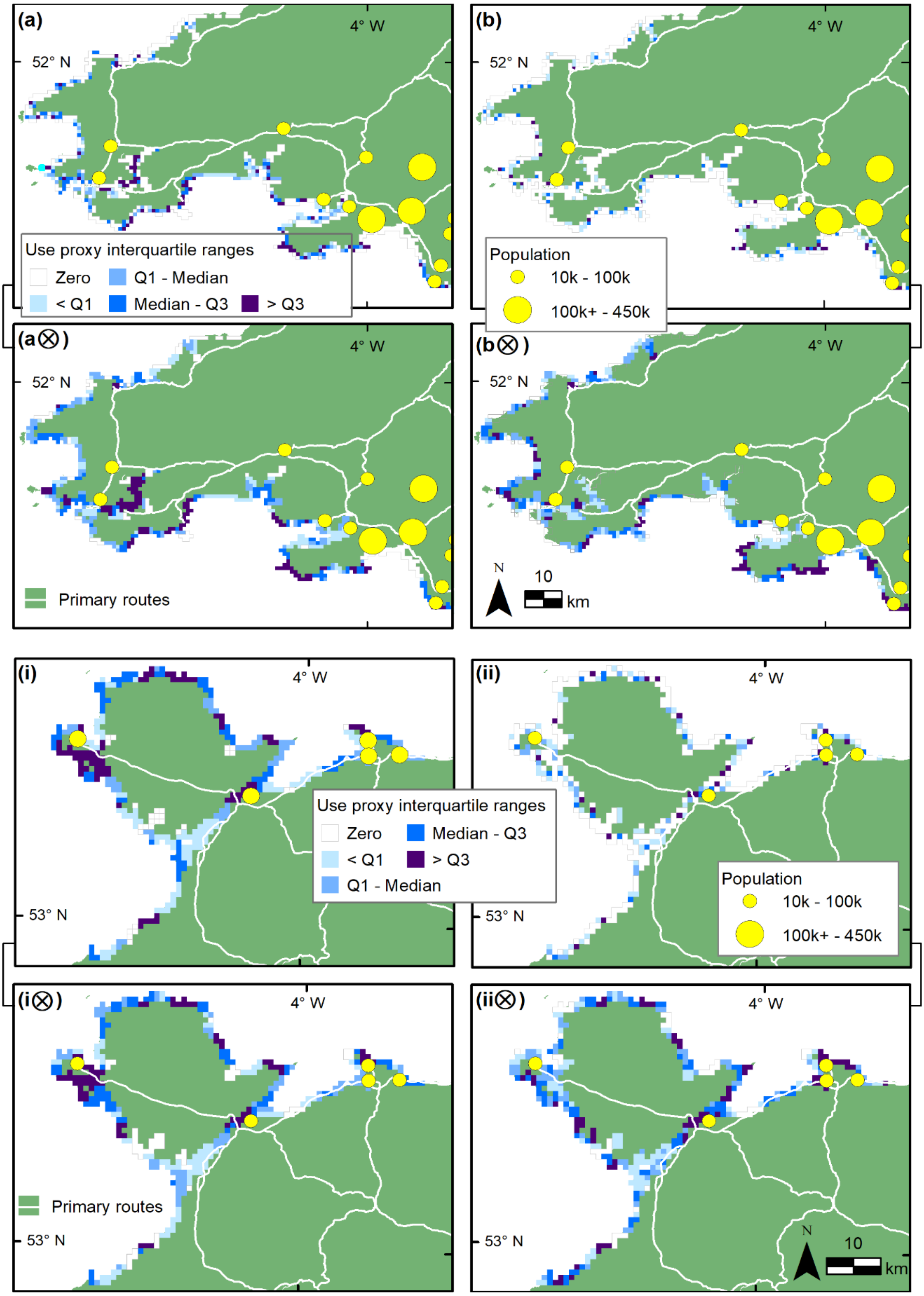

Figure 4-6. Extents of the Wales Activity Mapping (a & b, WAM) and FishMap Môn (i & ii, FMM) surveys showing quartiles of use proxy measures of marine recreational fishing activity at a resolution of 1 km^2. Left most maps are derived from the WAM and FMM surveys. The right most maps (b, ii) are derived from site densities mined for this article. A 3x3 box filter (alias, focal mean) was applied to rasters (a), (b), (i) and (ii) to derive filtered images indicated by ⊗.

Applying the box filter to the test and validation layers greatly improved the spatial distribution of intensity predictions. The convolution of course decreased the frequency of zeros across all layers by 40% ±9% S.D. and paired zeros by 85% and 79% for the FMM and WAM layers respectively. Nevertheless, correlations and inter-rater agreements were significant (Table 4-1) even after exclusion of paired zeros. The effect of applying the convolution on the spatial distribution of intensity can be visualized in the subfigure pairs of Figure 4-6 and heat map insets of Figure 4-7.

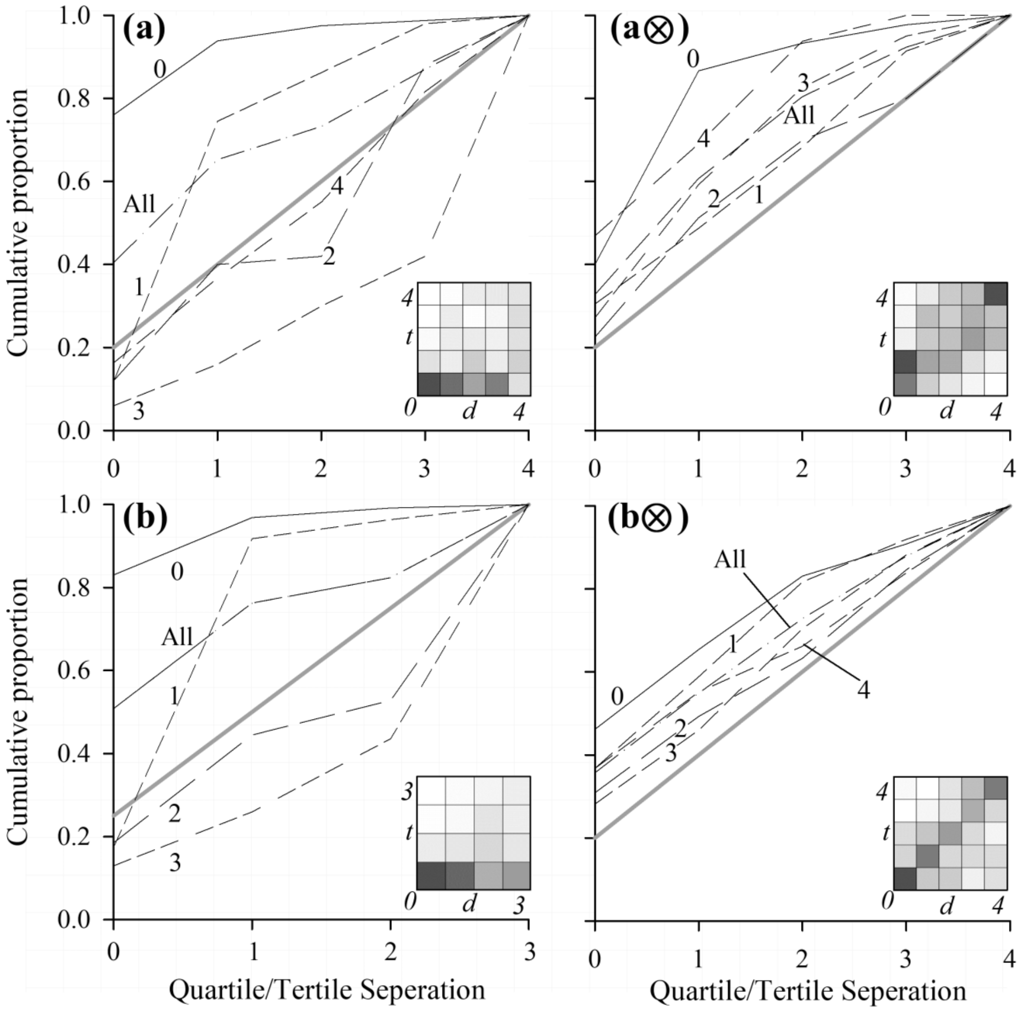

Figure 4-7. Cumulative proportion of quartile or tertile (Q|T) matches between FishMap Môn (Aron et al., 2014) and Wales Activity Mapping (Chambers et al., 2013) within the specified separation distance (x-axes). Insets are frequency distribution heat maps of quartile matches between the directed surveys (*d*) and this article (*t*). Quartiles were calculated from the FishMap Môn (a) and Wales Activity Mapping (b) surveys of marine recreational fishing intensity and are compared against Q|Ts from site densities derived from public and local knowledge. All data were calculated at a resolution of 1 km^2. Each line represents a Q|T (as labelled) of the directed survey and pertains to a heat map column. ⊗ Indicates Q|Ts of (a) and (b) after applying a 3x3 box filter to the original intensity data and recalculating Q|Ts. = expected cumulative proportion.

The convolution averages intensities within the filter kernel up to the diagonal maximum of 2.83 km. The effect is particularly noticeable in the point like data derived from local knowledge (i.e. the test data). However, the permutative inter-rater agreement and correlation tests prove that the observed agreement between the fuzzy test and fuzzy validation layers was not merely an intrinsic result of applying the convolution. The absence of activity was best predicted in the fuzzy data for both WAM and FMM layers, in addition low and high activity (i.e. quartiles 1 and 4) were better predicted than intermediate activity (i.e. quartiles 2 and 3). In fact, after convolution 87% of no activity in the test data are within 1 quartile of the validation zeros (Figure 4-7a⊗). Similarly for WAM, 65% of test data zeros are within 1 quartile and 83% are within 2 quartiles (Figure 4-7b⊗). For high activity (quartile 4), 69% (Figure 4-7a⊗) and 54% (Figure 4-7b⊗) are within one quartile of the validation data for FMM and WAM respectively.

4.5 DISCUSSION

Results show that heterogeneous public and local knowledge (LK) can be used as a proxy indicator of the qualitative levels of marine recreational fishing (MRF) activity across a large spatial area. The method only required limited human resources, and costs were minimal, yet mapped ordinal levels of shore activity at a resolution of 1 km^2 over 2,700 km of shoreline, 31% of which had no published data on the magnitude of activity. Results were validated against the comparatively resource intensive Welsh Government commissioned on-site interview based surveys FishMap Môn (abbr. FMM, Aron et al., 2014) and Wales Activity Mapping (abbr. WAM, Chambers et al., 2013), and found to be in fair agreement (Cohen's kappa, FMM, K = 0.36; WAM, K = 0.28). It should be borne in mind that FMM and WAM would also be imperfect assessments of the spatial distribution of activity.

4.5.1 COMPARING DATA SOURCES

Different sources of fisher knowledge were combined to predict sites favoured by fishers based on the frequency with which named sites occurred in public fisher knowledge, whereas previous research derived data from interviews or observations with participants (e.g. Close & Hall, 2006; Kafas et al., 2017; Léopold et al., 2014; Macdonald, Angus, Cleasby, & Marshall, 2014; Yates & Schoeman, 2013) or more rarely by exploiting single novel LK sources e.g. club records (Bennett et al., 1994; Gartside et al., 1999), magazines (Barbini et al., 2015; Elizabeth A Richardson et al., 2006), logbooks (Perzia, Battaglia, Consoli, Andaloro, & Romeo, 2016) and social media (Belhabib et al., 2016; D.R. Martin et al., 2012; Dustin R Martin et al., 2014; Shiffman et al., 2017). Conducting on-site interviews across the entire geographic extent of the study area would have been a major undertaking, but using heterogeneous LK sources achieved a resolution which matched that of FMM and site density was shown to be a fair proxy indicator of relative activity levels at the same spatial resolution. Standardization of venue density by shore length had at best no improvement in agreement with the validation data, thus shore length were not reliable predictors of the number of sites available to, or used by MRFs.

Comparatively large volumes of data were publicly available within the stated scope of the study. MRF in the United Kingdom is a popular recreational activity (Armstrong et al., 2013b), hence there is a rich seam of LK to mine. Additionally, over half of the sources which were used to identify sites were published on the World Wide Web. This approach is reliant on the accessibility and availability of LK and also the availability of site information from which to compile a gazetteer (e.g. United Kingdom Hydrographic Office, 2013). Thus, it could be argued that the application of the methodology is restricted to developed countries with popular MRF sectors who have a culture of sharing knowledge, however such knowledge may not be universal for all recreational fishing activity (Olsen & Thuen, 2013; Svensson, 2016), but see Belhabib et al. (2016). For example, despite private boat MRF days per annum accounting for ~30% of total annual trip days in England (Armstrong et al., 2013a), private boat angling was almost entirely absent from public LK. The increasing adoption of social media driven by smartphone use (G. Zhang, 2017) may increase opportunities to utilise digitally accessible LK, even in developing countries. However the switch towards social networking sites (e.g. Facebook) could be problematic as network effects (Katz & Shapiro, 1994) concentrates public LK to fewer private sites with user security options which prevent public access to user generated content.

4.5.2 Limitations

Shore MRF data is sparse over large areas and will contain many areas with no recorded activity. This was observed in these datasets and care must be taken in the interpretation, as zero activity areas can represent no data, zero activity, activity below the limits of detection, or simply be erroneous. In the context of coastal management, the absence of activity recorded at a site does not mean the site is not used or valued (personal observation, GGM in relation to the FMM spatial data). Variations in the seasonal patterns of species' distributions and long term species availability could affect the popularity and value of a site to MRFs, with some sites being utilized relatively intensively for short periods of the year in response to local abundances of valued species (GGM, personal observations).

In considering the validity of convolving any layers to improve the activity predictions derived from LK, it is important to consider why this improvement arises. It can be shown numerically that convolving two datasets drawn randomly from a Gaussian distribution will tend to increase the correlation. However, in this instance the permutation testing and observation of the maps showed this was not the explanation for the greatly increased agreement with the validation layers. In visually comparing the LK crisp layers with both crisp and fuzzy validation layers, there are many instances of elevated activity being in close proximity between layers, but not being precisely spatially coincident. This is probably the result of the mapping methods of FMM and WAM in which users drew areas used by MRFs and the point like data generated from the methodology in this study. If this spatial displacement occurs within a kernel distance of $\sqrt{2}$ kilometres (for the convolution used in this study) then the convolution will generally improve agreement with reality. However, many different convolutions could be applied (e.g. Gaussian filters, filters over different regions, different kernel shapes), and predicting the optimal convolution requires further investigation, probably using simulation approaches.

It is arguable whether applying the convolution would be a useful operation in creating MRF coastal use layers for consumption in GIS repositories to inform management processes. In the present study, the methodology of both directed surveys involved creating polygons which covered whole areas important to MRFs. Popular areas tend to be associated with secondary coastal features (e.g. sandy beaches) which are likely to cover more than one 1 km^2 cell and the convolution increases agreement in these instances. This was particularly apparent in south west Wales, e.g. on the north west and south east Gower peninsula, the Towy estuary and the area between Amroth and St. Govan’s Head (Figure 4-8). In areas where there are low numbers of small fishing platforms available (e.g. rocky headlands and jetties) then the likelihood of spatially erroneous predictions is increased however, it is suggested that review of sites using satellite imagery should be sufficient to identify accessible fishing platforms and to exclude inaccessible areas (GGM, personal observation). The extents of secondary coastal features, such as beaches and mudflats are also very easy to identify and combining these methods with manual review of sites and expert knowledge could provide sufficient evidence of use in lieu of randomised surveys. It would still be vital to engage with participants.

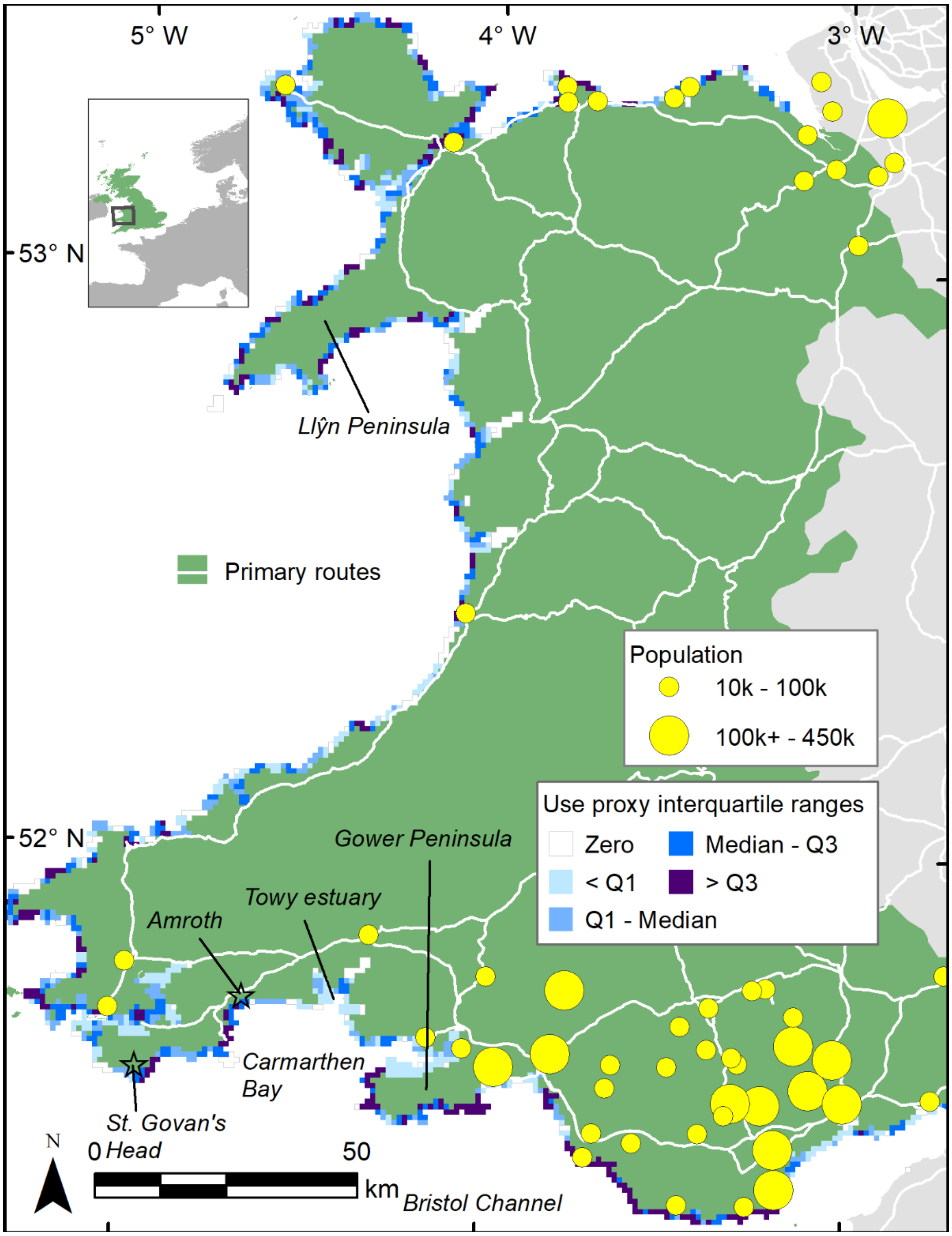

Figure 4-8. Interquartile ranges of 3x3 box filtered marine recreational fishing site densities across Wales, UK.

4.5.3 Planning

Shore-based MRF is typically open access in the UK with limited dependence on marine infrastructure and no compulsory reporting or licensing. This makes activity mapping substantially more difficult than many other non-recreational marine activities. The equitable management of marine resources is only possible when all competing interests and their socioeconomic benefits are known, and when their probable ecological effects are understood. Fishers can be protective of their knowledge (Maurstad, 2002; Olsen & Thuen, 2013; Svensson,

2016) and they are easily overlooked among the competing interest of higher profile human activities (K. St. Martin & Hall-Arber, 2008). Thus it is problematic to include consideration of the needs of MRFs, and the possible impacts arising from recreational fishing, when developing planning and management processes (Flannery, Healy, & Luna, 2018). It is implicit that a qualitative proxy for activity is a proxy for MRF site preferences, and using LK at the very least allows a spatial allocation of site importance to MRFs where survey data is unavailable or difficult to obtain. Multiple data sources can improve confidence and allow data to be cross validated, which will contribute to evidenced based approaches to conservation practice (Pullin & Knight, 2003).

Site density inter quartile ranges (IQR) were strongly correlated with the IQRs derived from directed surveys. To provide a numerical context, FMM had up to 2977 visits km^{-2} $year^{-1}$ (IQRs 21, 88, 230), and Goudge *et al*. (2009. 2010) recorded a maximum of 101 individuals at a site (IQRs 13, 18, 22). Conservation planning would benefit from mapping MRF activity hotspots by highlighting areas of potential conflict which are currently unknown or erroneously considered insignificant. MRF activity has been associated with major ecosystem changes, and removal of predatory fish has been shown to trigger trophic cascades (e.g. Dulvy et al., 2004; McClanahan, 1994). Altieri *et al*. (2012) offered evidence that recreational fishing contributed to die-off of north-western Atlantic salt marshes via a top-down trophic cascade. Altieri *et al*. (2012) reported a mean ±S.E of 5 ±1 fishers per site from visual observation. Impacts on protected species and habitats should be considered where high activity is suspected to occur because angling gear can harm non-target species (Asoh et al., 2004; Chiappone et al., 2005; Laist, 1997; Wells, Hofmann, & Moors, 1998). By combining ecosystems knowledge with potential impacts of recreational fishing and indicators of relative activity, research efforts into impacts and management efforts may be better directed.

Interactions with commercial fisheries can be important, for example the European Union imposed a total ban on recreational harvest of sea bass in ICES divisions IVb, IVc and VIIa – VIIk and multiple commercial technical measures. Conflicts can arise when commercial fisheries and recreational fisheries harvest the same stock, either because of spatial overlap in operational areas, or when mobile species are targeted (e.g. sea bass and Atlantic cod). Where fish stocks are compromised then unchecked MRF can impede stock recovery (Maggs et al., 2016; Sherwood & Grabowski, 2016). Under the precautionary principle (Garcia, 1994) harvest control measures (e.g. area closures) may be considered for recreational fisheries and scientific and expert knowledge can be sought where LK indicates that MRF activity is comparatively high and coincides with areas known to be important for the stock.

4.5.4 Concluding remarks

Gazetteer data and the local knowledge data are not randomly sampled nor are they independent. It would be incorrect to extrapolate activity proxies to produce population estimates of effort and catch to any wider population. Despite the demonstrated agreement with directed (but non-randomized) surveys, these data are strictly qualitative and are not a substitute for on-site surveys where demonstrably quantitative estimates on recreational activity are required (e.g. producing estimators of total monthly MRF effort). However, ordinal qualitative data does provide a comparative indicator of the value of spatial areas to MRFs and the level of activity by MRFs. These data provide layers which contribute to GIS management tools which are used in Marine Spatial Planning (e.g. Boyes et al., 2007) and can provide spatial information on MRF activity to conservation managers and marine policy decision makers. Having indicators of where fishers choose to fish is a necessary step in understanding the drivers behind MRF site selection and planning, but much work would remain to understand the effects of marine policy on MRF behaviour, which is necessary to implement effective management measures in response to changing environmental pressures.

Chapter 5

Accurate Estimation of Fish Length in Single Camera Photogrammetry with a Fiducial Marker

Chapter 5 has been submitted to the journal *ICES Journal of Marine Science* with some minor trimming of the text and is currently under major corrections.

https://academic.oup.com/icesjms

KH, MK and FV critically reviewed and revised the article. All other work was that of GGM.

No highlights were required by the journal.

5 Accurate Estimation of Fish Length in Single Camera Photogrammetry with a Fiducial Marker

5.1 ABSTRACT

Videogrammetry and photogrammetry are being used more widely in marine science for unsupervised data collection. The camera systems used to collect such data are complex. In contrast, digital cameras and smartphones are ubiquitous, convenient for the user and an image automatically captures much of the data normally recorded on paper as metadata. The limitations of such an approach are primarily attributed to the errors introduced through the

image acquisition process and errors introduced through lens distortion of the collected images. In the present study, a methodology is presented to achieve accurate total length estimates of fish without specialist equipment or proprietary software which could be used by any volunteer. Photographs of flat and fusiform fish were captured with an action camera using a (i) background fiducial marker, positioned at the distal plane of the subject, (ii) foreground fiducial marker, at the proximal plane of the subject and (iii) laser marker, projected on to the subject's surface. The intrinsic properties of the lens were modelled with OpenCV so images could be automatically undistorted. The accuracy of total length estimates were corrected for parallax effects using a novel iterative algorithm requiring only the initial length estimate and known morphometric relationships of the species. OpenCV was extremely effective in correcting image distortion, decreasing RMSE by 96% and the percentage mean bias error (%MBE) by 50%. Undistorting the image and correcting for parallax effects achieved the highest accuracy and also reduced estimation variance, achieving % MBE [95% CIs] of -0.6% [-1.0, -0.3] and reducing RMSE by 86% to 2.1%. Estimation of the lens subject distance using the similar triangles calibration method resulted in the best estimation of total length. The present study demonstrates that the morphometric measurement of different fish (or other) species can be accurately estimated with any camera and without expensive or bulky equipment.

5.2 INTRODUCTION

Cost reductions and advances in camera equipment and supporting technologies (e.g. durability, storage and computing capacity, connectivity, and supporting software) are stimulating research and development into the applications of photogrammetry and videogrammetry (hereafter referred to as photogrammetry) in marine research (Bicknell, Godley, Sheehan, Votier, & Witt, 2016; Struthers, Danylchuk, Wilson, & Cooke, 2015). Potential applications include remote electronic monitoring (virtual observation) of commercial fisheries to assess catch (e.g. White *et al.*, 2006; Chang *et al.*, 2010; Hold *et al.*, 2015; Bartholomew *et al.*, 2018) and bycatch (e.g. Pasco et al., 2009, Bartholomew et al., 2018), ecological studies using fixed cameras (e.g. Bouchet and Meeuwig, 2015; Schmid *et al.*, 2017), direct observational surveys (e.g. Harvey et al., 2001, Jaquet, 2006), behavioural studies (e.g. Nguyen et al., 2014, Claassens and Hodgson, 2018) and in aquaculture (e.g. Zion *et al.*, 2000; Costa *et al.*, 2006).

Length frequency data is particularly important in the assessment of fish stocks in recreational and commercial capture fisheries (Pauly & Morgan, 1987) however the collection of length measurements is time consuming and costly. Photogrammetry can increase throughput (Chang et al., 2010), mitigate against some biases (Faunce & Barbeaux, 2011; Harvey et al., 2001), and be cheaper per data point acquired than manual at-sea length sampling (Chang et al., 2010) and at-sea observation (National Oceanic and Atmospheric Administration, 2015b).

The equipment typically deployed for photogrammetry uses multiple parallel lasers (e.g. Deakos, 2010; Rogers *et al.*, 2017; Bartholomew *et al.*, 2018) or multiple cameras (e.g. Dunbrack, 2006; Rosen *et al.*, 2013; Neuswanger *et al.*, 2016). In parallel laser systems the lasers create a visible fiducial marker of known real-world length at the subject surface. When the plane of the subject surface is aligned with the plane of the camera sensor then the real-world length represented by an image pixel will be invariant across the subject provided the image and the subject are not distorted. Under these assumptions an accurate length estimate can be made. Multi-camera systems are mathematically more complex, but allow subject length (and other measures) to be estimated using triangulation methods (Hartley & Zisserman, 2004; Neuswanger et al., 2016). Accurate length estimates have also been derived by deployment of a simpler system by analysing images captured with a single camera and a physical fiducial marker of known length (Hold et al., 2015; van Helmond, Chen, & Poos, 2017).

Photogrammetry may widen the participation of non-scientists as novel sources of data. Citizen science projects are using smartphone applications to improve engagement with participants (reviews Hyder et al., 2015, Venturelli et al., 2017) and images are being used to identify species in images captured using smartphones (e.g. Fishbrain, 2018, Internation Game Fish Association, 2018). The assessment of marine recreational fisheries (MRF)—which can be data poor even in developed countries (ICES, 2017c)—may particularly benefit from the deployment of simple photogrammetry solutions. Surveys of MRF frequently have a diary phase in which anglers record details of their catch (ICES, 2014c). Volunteer based assessments may be the best means of collecting longitudinal data under budgetary limitations.

A huge number of historical images of fish exists in printed photographs and digitally accessible knowledge (DAK) repositories (e.g. social media). Single photographs and opportunistic fiducial markers (i.e. an object of known real-world size being present in the image by chance) have been used to investigate long term temporal population structure changes in fish (Canese & Bava, 2015; McClenachan, 2009) and several papers have used DAK in the form of photographs or videos to research marine recreational fisheries to, describe the fishery (Giovos et al., 2018), to investigate illegal practice (Shiffman et al., 2017), to assess disease (Rizgalla et al., 2017), and to estimate size using opportunistic fiducial markers (Belhabib et al., 2016).

The existing use of fiducial markers to estimate length has focused on image and video capture using comparatively complex and costly multi-camera or multi-laser systems. Single-camera photogrammetry using a physical fiducial marker or lasers have limited error by controlling the camera model, the lens-subject distance or the framing of the subject within the camera’s field of view (e.g. Hold et al., 2015, Rogers et al., 2017). These approaches are impractical to deploy in large scale volunteer and citizen science projects, or to fisheries in severely resource limited countries. However, smartphone ownership is high in developing countries (Median ~37%, Poushter, 2016) and adoption is expected to increase (Poushter, 2016; G. Zhang, 2017).

To accurately estimate length from images using a fiducial marker several corrections are necessary. Cameras have different intrinsic tangential distortion, where the sensor plane is not perpendicular to the optical axis. Additionally, the wide-angle lenses typical of action cameras and smartphones exhibit radial distortion. These factors introduce systematic length estimation errors as the real-world length represented by pixels across the captured image plane varies with the location of the pixel in the image. Any estimation of real-world size can be biased by the changing depth and pose between the subject, the fiducial marker and the camera (*parallax effect*). Camera calibration is well understood (see Szeliski, 2010 pp288-295) nonetheless, correcting fiducial marker-made length estimates for subject pose in single camera systems has received little attention. Corrections can be made using the thin-lens equation provided the lens-subject distance is known. However, measuring the lens-subject distance is impractical for some uses, e.g. in volunteer based projects where the volunteer could not be expected to accurately measure the lens-subject distance each time an image was captured.

This article aims to introduce a methodology to minimise errors in morphometric measurements of fish (and other organisms) when using single camera photogrammetry. The methodology is particularly relevant to the automation of length extractions in machine vision pipelines for volunteer led applications used in the assessment of recreational fisheries or small scale and developing artisanal or commercial fisheries. The emphasis is placed on methods to reduce length estimation biases when deploying a foreground fiducial marker. Length estimation using a foreground fiducial marker has received little attention, yet offers several advantages when deployed in volunteer led applications, including very low cost, high portability and size estimates cannot be increased by moving the subject closer to the camera. For context the more commonly deployed paired laser and background fiducial markers are also included.

The objectives are to (i) empirically compare the accuracy of low-cost foreground, background and laser fiducial markers; (ii) validate the effectivity of using the open source, platform-agnostic OpenCV API in correcting intrinsic lens distortion in any camera; (iii) describe methods to minimise error in length estimates made with fiducial markers; and (iv) empirically compare the effectiveness of applying a lens distortion correction and parallax correction without prior knowledge of the lens-subject distance.

5.3 METHODS

5.3.1 IMAGE ACQUISITION AND *ACTUAL TL* MEASUREMENT

Photographs of European sea bass (*Dicentrarchus labrax*, n = 43) were gathered *in-situ* at a commercial fish processor. Images of common dab (*Limanda limanda*, n = 32) were taken in the laboratory. The camera system was a Nextbase 512G camera encased in a custom housing with 12v battery. The Nextbase 512G optical system has a wide angled field of view (FOV) and significant barrel distortion, which allowed the effectiveness of lens distortion correction to be evaluated. The camera housing was mounted on a Manfrotto 244 variable friction arm and bracket. Projective distortion was minimised by using spirit levels to ensure the principle lens plane and the surface on which the photographic subject lay were parallel.

The camera recorded in video mode at a 1280 × 720 pixel resolution so images were captured without perturbing the camera. Frames were manually extracted from the captured video. The distance between the surface on which the subject lay and the front glass of the lens housing was measured with a 1 meter steel rule (required for depth adjustment as outlined later) and the total length (TL) of the subject was measured using a fish measuring board (henceforth *actual TL*, **l**) with the caudal lobes pushed gently together and then allowed to settle without further coercion. All measuring rules were validated with an EC class 1 certified rule. Throughout, all lengths refer to the TL unless otherwise specified.

The precision and accuracy of three types of fiducial markers were compared (Figure 5-1), these were; (i) marker positioned at the backplane (*distal plane*) of the subject (henceforth *background marker*); (ii) paired lasers projected onto the near surface of the subject (henceforth *laser marker*); and (iii) a marker positioned on the subject surface (*proximal plane*) closest to—and parallel with—the plane of the camera lens (henceforth *foreground marker*).

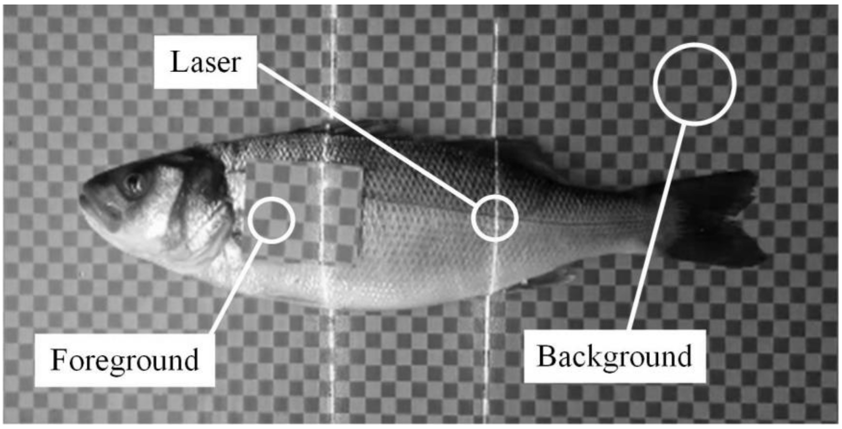

Figure 5-1. Illustrated position of the three fiducial markers with respect to the subject. Cells are $1cm^2$. The background fiducial marker is positioned at the distal plane of the fish, parallel with the camera sensor. The foreground fiducial marker is at the proximal plane of the fish. The two laser markers are approximately at the proximal plane and an intermediate position between the proximal plane and the distal plane.

The foreground and background markers were a chessboard of $1cm^2$ cells printed on waterproof vinyl and mounted on a polycarbonate sheet (Figure 5-1). The laser marker used two parallel-paired lasers (Odiforce, 3-5mW Green Laser Module) mounted in the camera housing. The distance between the laser lines at the centre of the background marker was recorded because the laser lines were not parallel at the scale of interest due to fabrication errors.

Image frames were extracted from the video and real subject TL estimates made in ImageJ (Schneider, Rasband, & Eliceiri, 2012) for each of the 3 fiducial markers. The background fiducial marker real-world length per pixel (RWLPP) was calculated across the whole length of the fish. The pixel length of the fish was measured in the image by the line segment joining the tip of the snout through the centre of the caudal peduncle and the fork to the intersection with the imaginary line between the tips of the caudal fin.

5.3.2 HIERARCHY OF LENGTH CORRECTION REFINEMENTS

Figure 5-2 illustrates the position of each fiducial marker type in relation to the camera. It is apparent that the estimation of the RWLPP is dependent on the distance between the fiducial marker and the camera. Errors in TL estimation arise because of variation in the distance between the fiducial markers and the subject profile. Errors are also caused by image distortion arising from the intrinsic properties of the camera-lens system (henceforth *intrinsic camera properties*). It is evident that these two sources of error need to be corrected to produce accurate TL estimates.

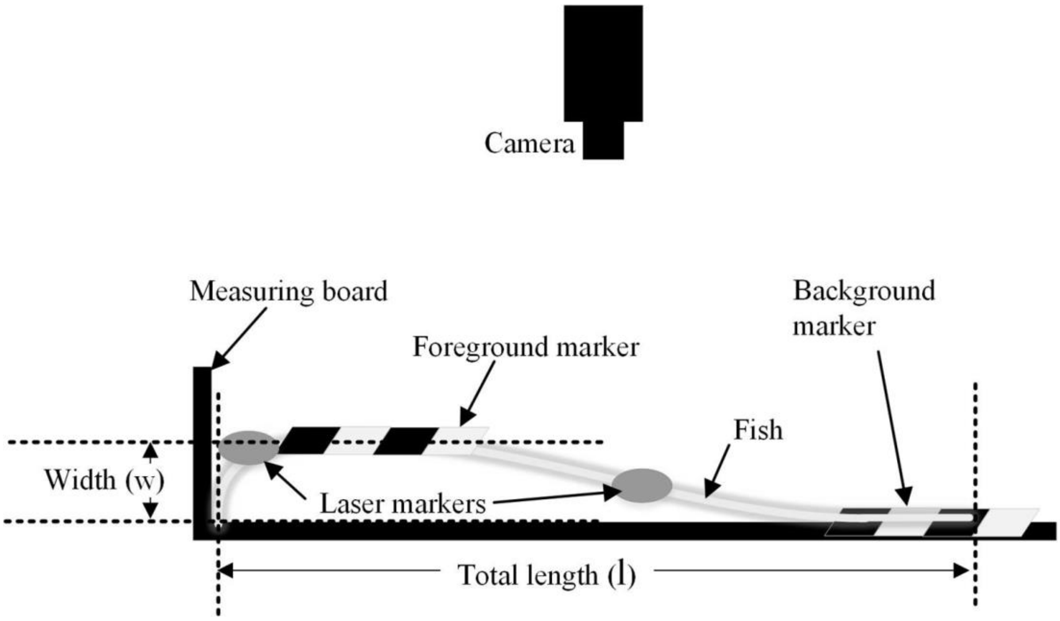

Figure 5-2.The measurement of total length with a fish board, and the relational position of foreground, background and laser fiducial markers and the camera. Width is used here to describe the elevation of the subject above its distal plane, as in fish with a fusiform morphology.

5.3.3 CORRECTING FOR IMAGE DISTORTION

To correct for tangential distortion, radial distortion and lens misalignment, the intrinsic parameters of the camera at a fixed zoom (focal length) and the lens distortion coefficients need to be calculated (Szeliski, 2010). Multiple images of a regular 2D pattern were captured in different orientations and the intrinsic camera matrix and distortion coefficients calculated using Python 3.5 and OpenCV (OpenCV team, 2018). This camera profile is saved and can then be reused to undistort images taken with the same camera for a given focal length. Appendix G lists the code used for camera profile creation and undistorting images.

The efficacy of the distortion correction was estimated by photographing a chessboard pattern and manually marking the vertices both before and after distortion correction. On an image without radial distortion, points should lie on straight lines, so the x and y coordinates were regressed and the residuals used to calculate the Euclidean distance in pixels of the marked point from the idealised vertex. The root mean squared error (RMSE) was calculated.

Errors arising from differences in pose were minimised by ensuring the camera and subject were aligned as previously described and by ensuring the subject was placed on the background marker with minimal body distortion. Henceforth TL estimates taken from an undistorted image are known as *undistorted TL* (l_{und}).

5.3.4 Correcting for subject profile

In the case of the laser and foreground markers, the width of a fusiform fish (w, Figure 5-3) causes an underestimate of the RWLPP, therefore TL (l, Figure 5-3) is also an underestimate. TL estimations made with a foreground marker can be corrected using a well-known manipulation of the thin lens equation where $a = b(1 - w)/d$ (Figure 5-3).

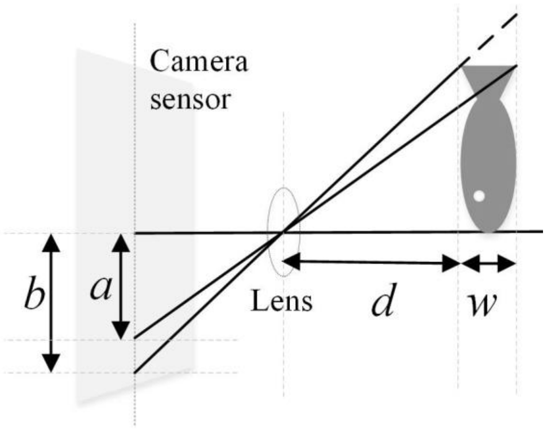

Figure 5-3. The thin-lens model, which relates real-world lengths to image formation at the camera sensor.

This correction (henceforth *depth corrected TL*, l_d) can be interpreted as adjusting the RWLPP to be the same as the RWLPP if the foreground fiducial marker was positioned at the distal plane. In calculating l_d, the width of the fish is required, which can be estimated from the length provided the length-width relationship is known. *Depth corrected TL* is subject to a systematic error because the estimated length used to derive w is itself an underestimate. The length was corrected iteratively according to the process in Figure 5-4. This correction is reported as *iterative corrected TL* (l_{iter}), which is $l_{und} + l_{cor}$ where l_{cor} is the sum of iteratively calculated lengths larger than the minimum threshold of 0.1 mm (See Appendix I line 37 for stop criteria detail).

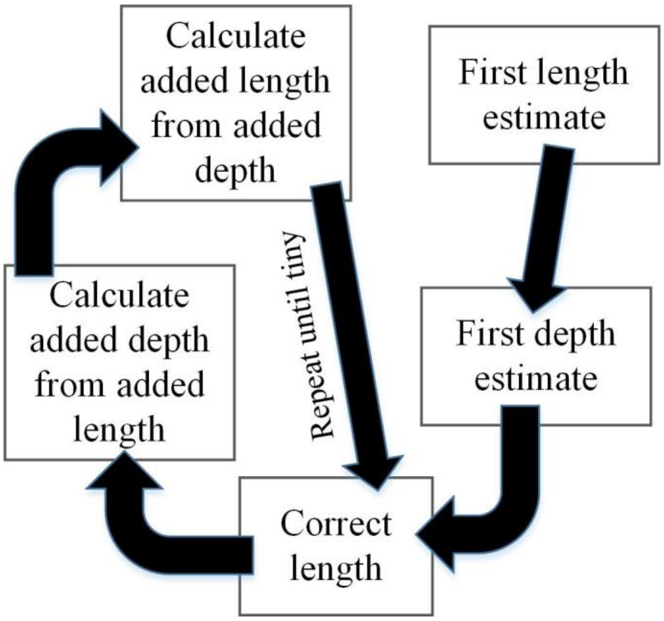

Figure 5-4. Iterative process to improve accuracy of total length estimates. The iteration was repeated until the incremental length to be added at was less than 0.1 mm.

Iterative corrected TL does not account for the varying width profile of the fish between the proximal and distal subject planes. To test this correction methodology, we compared the dab with sea bass and calculated the mean width (mm) for each species. No length-width morphometric data were available for dab hence the mean width was calculated from the samples. The mean widths were measured by dividing fish samples through the long axis of the coronal plane. The bisected samples were then photographed against a white background. Images were threshholded (i.e. subject pixels set to white, background pixels set to black), then the *standardised mean width* $\hat{w}$ was derived from the mean pixel width across the threshholded images (Figure 5-5), according to $\hat{w} = \left(\frac{1}{n} \cdot \sum_1^n w_i\right) / \max(w_i)$ where *n* is the number of pixel columns and w_i is the height in pixels of the i^{th} column. This factor was used to correct the *iterative corrected TL* to derive the *profile corrected TL* (l_p) according to $l_p = \hat{w} \cdot l_{cor} + l_{und}$. Appendix H lists the code to calculate the mean width from a threshholded image.

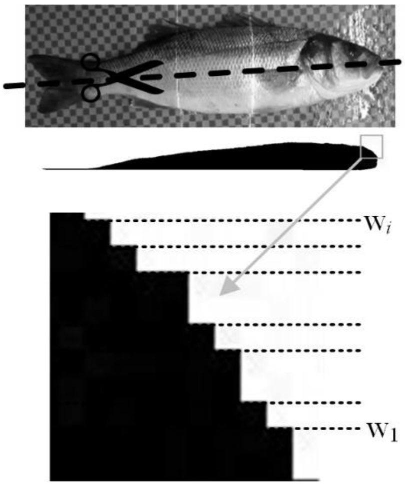

Figure 5-5. Fish samples were cut through the coronal plane (top), then the width profile photographed against a white background.

The image was then threshholded (middle) and the mean pixel width calculated (bottom) using Python and OpenCV, to give an estimate of the mean width.

5.3.5 LENGTH ESTIMATION WITH UNKNOWN LENS-CAMERA DISTANCE

Length corrections require prior knowledge of the distance between the lens and the subject (*d* in Figure 5-3). This is impractical for many applications. Two methods based on the thin lens model can be used to estimate *d* if a fiducial marker appears in the image. Firstly, *d* can be estimated if we know some properties of the camera, according to $d = (f \cdot h \cdot \acute{s})/(\acute{h} \cdot s)$ where *f* is the focal length, *h* is the real-world size of the fiducial marker, $\acute{s}$ is the sensor height in pixels, $\acute{h}$ is the height of the fiducial marker in pixels and *s* is the real-world size of the sensor. Secondly, *d* can be estimated by taking one (or more) calibration images with a marker of known length, according to $d = (h_c \cdot h \cdot f)/(\acute{h_c} \cdot \acute{h})$, where f is the focal length, h_c is the real-world size of the calibration marker, h is the real-world size of the fiducial marker, $\acute{h_c}$ is the height in pixels of the calibration marker and $\acute{h}$ is the height of the fiducial marker in pixels.

Both methods were used to estimate *d* (Figure 5-3) in the calculation of the *profile corrected TL* and are reported as *calibrated profile corrected TL* and *sensor profile corrected TL*. Appendix I lists the core functions used to produce these corrections. Both methods were used to estimate d in the calculation of the profile corrected TL and are reported as calibrated profile corrected TL and sensor profile corrected TL. Supplementary materials C lists the core functions used to produce these corrections. In summary, the mean bias error (MBE) is reported for the variables listed in Table 5-1. Mean bias error (MBE) is calculated according to $MBE = 1/n \cdot \sum_{i=1}^{n} \hat{y_i} - y_i$, where ŷi is the ith estimate of the actual TL yi. Percent MBE is given by $\%MBE = 100/n \cdot \sum_{i=1}^{n} \hat{y_i} - y_i/y_i$.

Table 5-1. Description of variables used in the present study. RWLPP, real world length per pixel, i.e. the number of millimetres a pixel in the image represents (units of mm pixel^{-1}). Marker, the fiducial marker. TL, total length. The width calculation (*w*) was parameterised from Poli et al. (2001). See Figure 5-3 for the parameters *a*, *b*, *d* and *w*.

Variable	Derived From	Description	Additional Detail
Actual TL, l	N/A	TL measured using a fish board.	Physical measurement taken and recorded by a person.
Distorted TL, l_{dis}	Distorted image	TL estimated from an image without any correction for lens distortion by manual measurements in ImageJ.	RWLPP is the *real-world marker length / marker length in pixels* in the native image. $l_{und} = RWLPP \cdot Fish\ TL\ in\ pixels$
Undistorted TL, l_{und}	Undistorted image	TL estimated from an undistorted image, reported for all three fiducial marker types.	As *Distorted TL*, but images were undistorted using the lens profile of the camera created in OpenCV.
Depth corrected TL, l_d	Undistorted TL	Adjustment for the difference in the distance between the proximal and distal plane of the subject. Not applicable for the background marker. Uses the actual lens subject distance in the calculation.	l_d is calculated by re-estimating RWLPP ($\widehat{RWLPP}$) using an estimate of the width of the fish w, where $w = 0.136 \cdot l_{und} - 0.367$. Given *d* is the distance between the lens and the fiducial marker then, $\widehat{RWLPP} = marker\ length\ pixels \cdot (1 - w)/d \cdot real\text{-}world\ marker\ length$
Iterative corrected TL	Depth corrected TL	Apply an adjustment for the initial underestimate of TL.	l_{und} is an underestimate of *l*, hence *w* is also an underestimate. The length adjustment $l_d - l_{und}$ is taken and the corresponding increase in w, Δw is calculated. The change in $\widehat{RWLPP}$ as above using Δw and the new total length recalculated. This process is repeated until the length added falls below a threshold of 0.1 mm.
Profile corrected TL	Iterative corrected TL	Apply an adjustment accounting for the mean profile width of the subject, i.e. correcting for the parallax effect.	The standardised mean width, $\widehat{w}$ of the fish was calculated and this factor was used to correct the *iterative corrected TL* to derive the *profile corrected TL* (l_p) according to $l_p = \widehat{w} \cdot l_{cor} + l_{und}$.
Calibrated profile corrected TL	Depth corrected TL	Recalculates depth corrected TL using an estimate of the lens-subject distance using similar triangles, then	Replaces *d* in the calculation of $\widehat{RWLPP}$, with an estimate of *d* according to $d = (f \cdot h \cdot \acute{s})/(\acute{h} \cdot s)$ where *f* is the focal length, *h* is the real-world size of the fiducial marker, $\acute{s}$ is the sensor height in

		applies the same process used to calculate the profile corrected TL.	pixels, $\acute{h}$ is the height of the fiducial marker in pixels and s is the real-world size of the sensor.
Sensor profile corrected TL	Depth corrected TL	Recalculates the depth corrected TL based on the thin lens equation parameterised with camera properties, then applies the same process used to calculate the profile corrected TL.	Replaces d in the calculation of $R\widehat{WLP}P$ where d is estimated by taking a calibration image with fiducial marker and calculated according to $d = (h_c \cdot h \cdot f)/(\acute{h_c} \cdot \acute{h})$, where f is the focal length, h_c is the real-world size of the calibration marker, h is the real-world size of the fiducial marker, $\acute{h_c}$ is the height in pixels of the calibration marker and $\acute{h}$ is the height of the fiducial marker in pixels.

Homogeneity of variance was determined using Levene's test. Where data were heterogeneous, estimators of central tendency were calculated using a 1000 sample bias corrected accelerated bootstrap (BCA) in SPSS (IBM Corp, 2011). A weighted least squares general linear mixed model (wls-GLMM) was used to compare percent errors ($e_{\%}$) for the species (random) and marker (fixed) factors. The vector of weights (W) were calculated as follows. Let $|R|$ be the vector of absolute non-standardized residuals from the regression $e_{\%} \sim species + marker + \ species * marker$. Then let P be the vector of predicted values of $|R| \sim species + marker + species * marker$. Then the vector of weights W, is $W = 1/P^2$.

5.4 RESULTS

Measured sea bass sizes ranged between 279 mm and 580 mm, and dab sizes were between 100 mm to 282 mm (Figure 5-6). For bass, the length-width relationship was taken from data published in Poli et al. (2001) to give $width = 0.136 \cdot total\ length - 0.367$, where length is measured in centimetres. For dab (n = 21) $width = 0.087 \cdot total\ length - 2.915$ (R = 0.98, p < 0.001). The mean widths were estimated as 0.598 and 0.505 for sea bass and dab respectively.

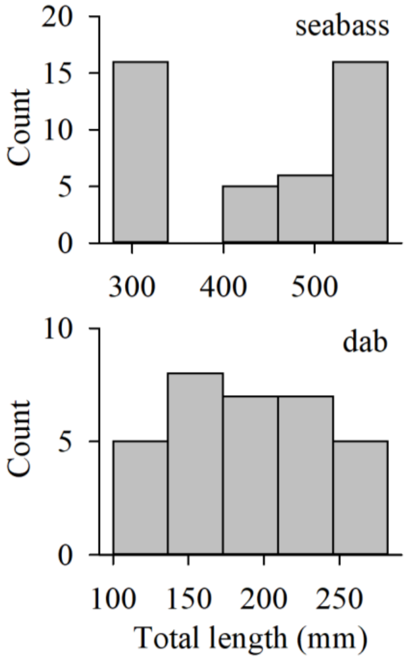

Figure 5-6. Actual total length histograms for dab and sea bass.

5.4.1 DISTORTION CORRECTION

OpenCV (OpenCV team, 2018) was successful in reducing the radial distortion of the optical system of the NextGen 512G camera (Figure 5-7). In captured images a pixel represented distances of between 0.5 mm and 1.0 mm. The absolute deviation of the vertices from idealised straight lines in distorted images was mean ±S.D = 18.2 pixels ±11.3 compared to a mean ±S.D of 0.7 pixels ±0.4 for the undistorted images and the RMSE was reduced by 96% (distorted RMSE = 21.4 pixels, undistorted RMSE = 0.76 pixels).

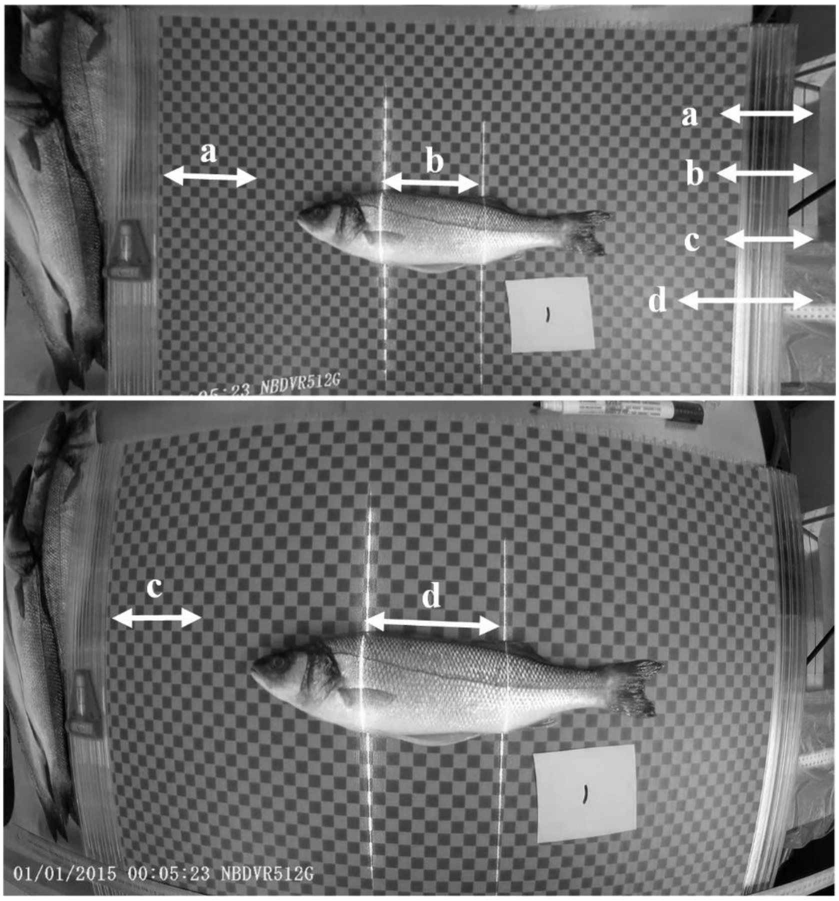

Figure 5-7.Example image of sea bass, with noticeable radial distortion (bottom) and after the image has been undistorted using OpenCV in Python (top). The lines a, b, c and d were all set to a length of 10 cm against the background fiducial marker.

5.4.2 DISTORTION CORRECTED LENGTH ESTIMATES

The laser and foreground fiducial markers substantially underestimated *actual TL* without any lens correction for laser and foreground markers and in both species (Figure 5-8; Table 5-2; aggregated %MBE [95% CIs] = -12.9% [-14.1, -11.7]) and this bias was still substantial for both markers for *undistorted TL* (Figure 5-8; Table 5-2; aggregated %MBE [95% CIs], -6.5% [-7.1, -5.9). Estimations made using the background marker were accurate, precise and robust to lens distortion, but overestimated TL in both species (Figure 5-8; Table 5-2; aggregated %MBE [95% CIs], 2.4% [2.1, 2.7]) and undistorting the images improved background MBE by just 0.6 mm for bass and 0.7 mm for dab (aggregated %MBE [95% CIs], 2.3% [1.9, 2.7]).

The magnitude of the absolute error increased linearly for all marker types with significant non-zero gradients (BCA bootstrap linear regression, $p < 0.05$). It is apparent that the systematic linear increase in error was most marked in estimates of dab TL and with the laser marker (Figure 5-8).

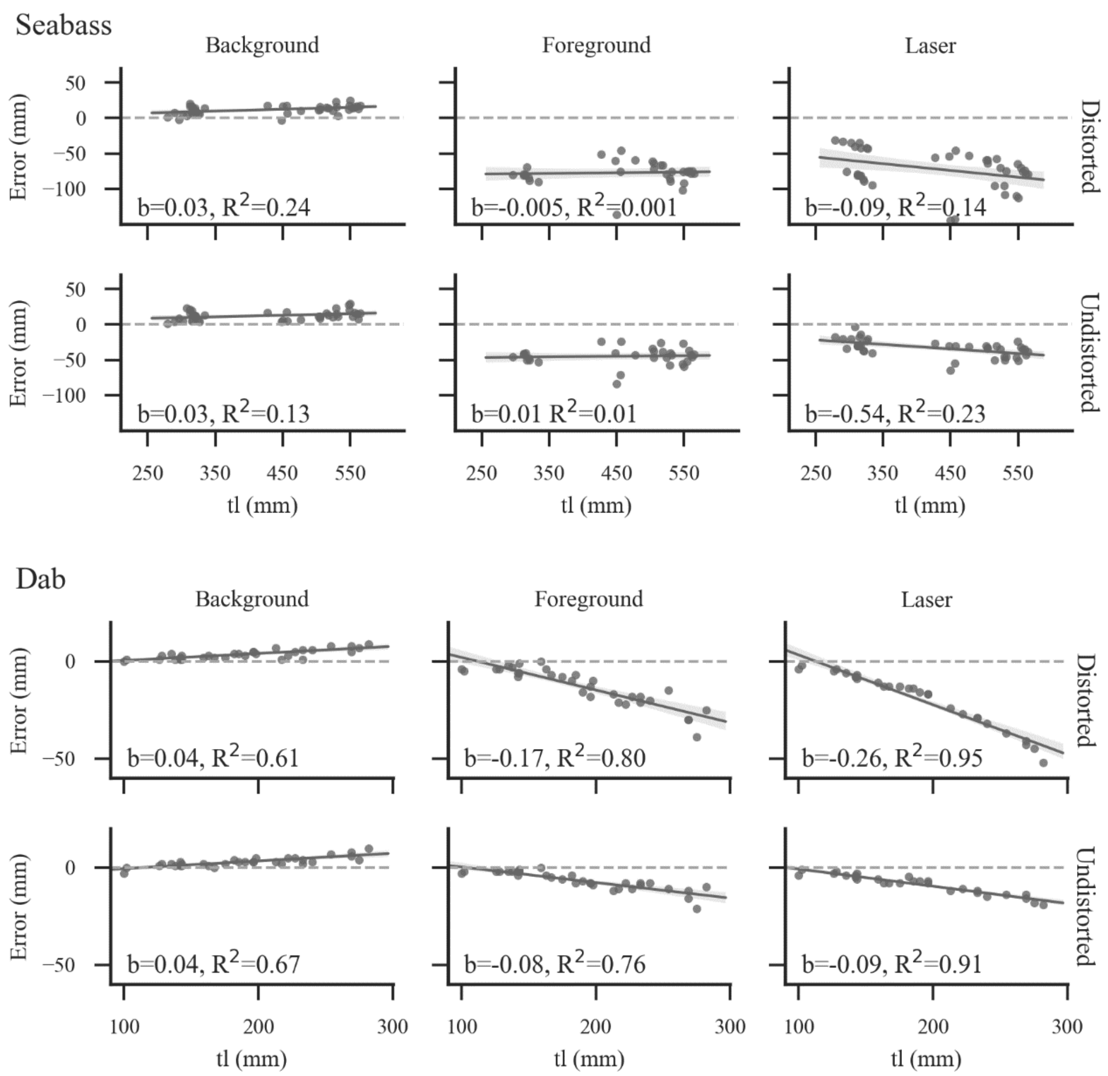

Figure 5-8. Error in estimation of total length (TL) for sea bass and common dab , using foreground, background and laser fiducial markers from images without any correction for camera-intrinsic radial and tangential distortion (distorted) and after correcting images for intrinsic distortion (undistorted). Plot is *actual TL* measured using a fish board *vs*. (corrected total length - *actual TL*), hence a negative error represents an underestimate of TL. Shaded line is the 95% confidence intervals.

5.4.3 Length estimate refinements

Applying successive width profile corrections to images substantially improved accuracy for both species when compared to *undistorted TL* (Figure 5-9) in both species with an overall reduction in %MBE of 95% (i.e. -12.9% to -0.6%). *Profile corrected TL* had the greatest accuracy and lowest variance (Figure 5-9, Table 5-2) with an aggregated mean %MBE [95% CIs] of -0.6% [-1.0, -0.3] and RMSE was reduced by 86% from 14.8% to 2.1%. In both species, *profile corrected TL*, *calibrated profile corrected TL* and *sensor profile corrected TL* tended to suppress error scaling with increasing TL when compared against non-profile based corrections. This effect is indicated by the reduced magnitude of the linear regression coefficients across the various profile corrections (Figure 5-9; ANOVA, $F_{(1, 28)}$, = 6.26, p = 0.02, η^2 = 0.19).

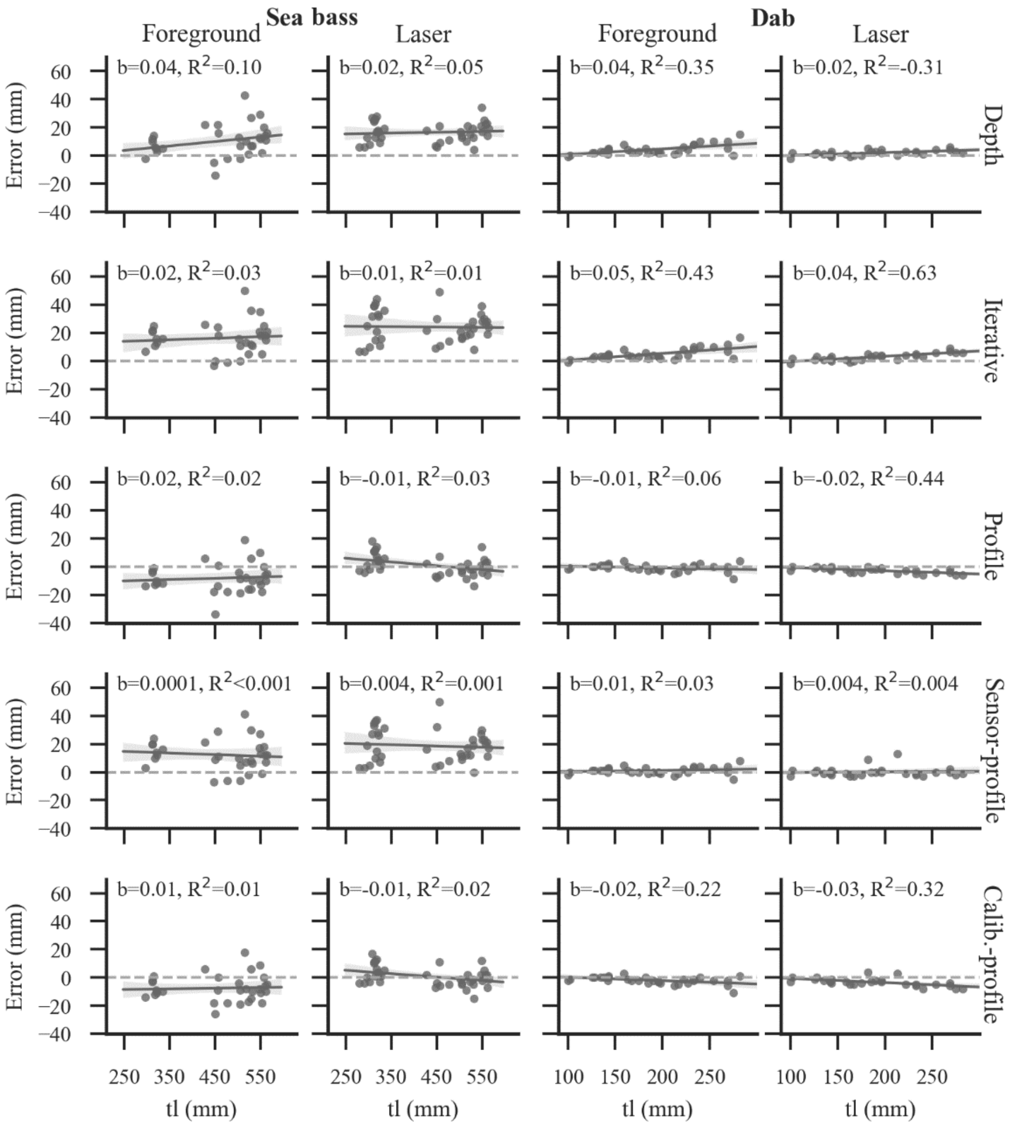

Figure 5-9. Errors in estimation of total length (TL) for European sea bass and common dab, using foreground and laser fiducial markers after correcting images for radial and tangential distortion. Plot is *actual TL* measured using a fish board *vs.* (corrected total length - *actual TL).* A negative error represents an underestimate of TL. Linear regression coefficient (*b*) and R^2 reported. Note that the null model (i.e. $b = 0$) indicates that error was untrended. 95% bootstrapped (n = 1000) confidence intervals appear in grey.

Table 5-2*. Mean bias errors (MBE) ± standard deviation (S.D.) and 95 % confidence intervals (CIs) for 7 different total length estimates made using background, foreground and laser fiducial markers from photographs of European sea bass and common dab.

	Background				Foreground				Laser			
	n	**MBE±S.D.**	**Range**	**95% CIs**	**n**	**MBE±S.D.**	**Range**	**95% CIs**	**n**	**MBE±S.D.**	**Range**	**95% CIs**
European sea bass (*Dicentrarchus labrax*)												
Uncorrected	43	12.8 ±6.9	-3 – 27	10.9 – 14.7	35	-78.5 ±16.3	-136 – -46	-83.8 – -73.5	43	-73.2 ±26.8	-145 – -31	-81.6 – -65.2
Undistorted	43	13.3 ±7.9	1 – 39	11.1 – 15.6	35	-44.7 ±13.2	-84 – -23	-48.9 – -40.5	43	-33.3 ±12.3	-65 – -3	-36.9 – -29.8
Depth	-	-	-	-	35	11.4 ±11.7	-14 – 43	7.6 – 15.4	43	18.0 ±10.0	4 – 55	15.3 – 21.1
Iterative.	-	-	-	-	35	17.5 ±12.0	-3 – 50	13.6 – 21.7	43	25.8 ±13.2	7 – 69	22.2 – 29.9
Profile	-	-	-	-	35	-7.5 ±10.9	-34 – 19	-10.8 – -4.2	43	2.0 ±9.0	-14 – 31	-0.4 – 4.8
Sensor profile	-	-	-	-	35	13.2 ±11.3	-7 – 41	9.5 – 17.1	43	20.0 ±13.1	0 – 58	16.6 – 23.6
Calib profile	-	-	-	-	35	-6.7 ±10.0	-26 – 18	-9.8 – -3.5	43	1.7 ±8.8	-15 – 30	-0.8 – 4.6
Common dab (*Limanda limanda*)												
Uncorrected	32	3.7 ±2.4	0 – 9	2.9 – 4.5	32	-12.9 ±9.7	-39 – 0	-16.0 – -10.3	28	-19.3 ±14.2	-52 – -2	-24.4 – -15.0
Undistorted	32	3.0 ±2.5	-3 – 10	2.2 – 3.8	32	-6.8 ±4.8	-21 – 0	-8.5 – -5.3	28	-8.6 ±5.0	-19 – -1	-10.3 – -7.1
Depth	-	-	-	-	32	4.3 ±3.5	-1 – 15	3.2 – 5.4	28	1.9 ±1.9	-2 – 6	1.2 – 2.5
Iterative	-	-	-	-	32	4.9 ±3.9	-1 – 17	3.8 – 6.1	28	2.9 ±2.6	-2 – 9	2.1 – 3.8
Profile	-	-	-	-	32	-0.8 ±2.6	-9 – 4	-1.9 – .1	28	-2.7 ±1.9	-6 – 0	-3.4 – -2.1
Sensor profile	-	-	-	-	32	1.1 ±2.5	-5 – 8	0.2 – 2.1	28	0.1 ±3.4	-3 – 13	-1.1 – 1.6
Calib profile	-	-	-	-	32	-1.9 ±2.6	-11 – 3	-3.0 – -1.0	28	-3.2 ±2.9	-8 – 4	-4.3 – -2.2

*All numbers represent millimetres.

Of the two profile corrections which used an estimation of subject-lens distance, *calibrated profile corrected TL* was more accurate (*Calibrated profile corrected TL*, %MBE [95% CIs], M = -0.8% [-1.1, -0.4]; *Sensor profile corrected TL* %MBE, M = -2.4% [-1.8, -2.9]) and consistent (RMSE; *calibrated profile corrected TL*, 2.1%; *sensor profile corrected TL*, 3.9%).

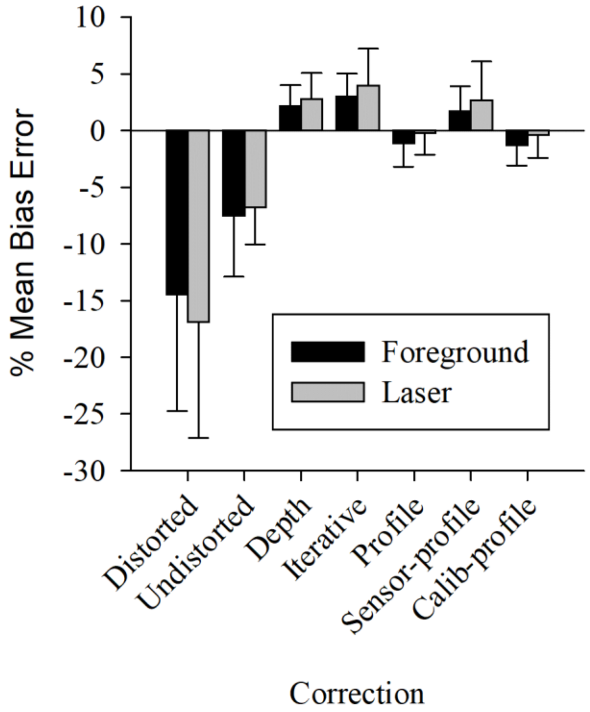

Figure 5-10. Species combined percent mean bias errors of estimated total length from a background, foreground and laser fiducial markers with standard deviation.

Profile corrected TL %MBE was marginally reduced in the laser marker compared to the foreground marker (Figure 5-10; mean [95% CIs]; laser, -0.18% [-0.6, 0.3]; foreground, -1.1% [-1.6, -0.6]), but this was not a significant reduction in bias (wls-GLMM, $P > 0.05$). Error in estimating TL by species was also not significant (wls-GLMM, $P > 0.05$).

5.5 Discussion

By applying corrections for intrinsic lens distortion, the accuracy of length estimates using foreground and parallel laser fiducial markers was significantly improved in both of the test species. Accuracy was further enhanced by applying increasingly refined corrections to account for the changing distance between the camera and the subject across the surface of the subject. The best accuracy was achieved (%MBE = -0.6 %) when the lens-subject distance was manually measured to millimetre accuracy and the total length estimated by iteratively accumulating additional lengths and adding the accumulated sum to the initial length estimate. This error was not significantly different from a %MBE of zero and at -0.6% represents approximately 2 cm per metre. This error magnitude is comparable to that observed by Hold *et al.* (2015) who used a fiducial marker to estimate carapace size in crab (*Cancer pagurus, 0.1 %*) and lobster (*Homarus gammarus, 0.6 %*). Similar error magnitudes have also been observed when using paired lasers (e.g. Deakos 2010, 0.4 % in *Manta alfredi*) or multi-camera systems (e.g. Rosen *et al*. 2013 1.0 % in *Scomber scombrus*, *Pollachius virens* and *Pollachius pollachius*). Population studies that examine size-structure typically bin estimates into size classes hence this level of error should not unduly bias biomass removal estimates, size selectivity, trends in trophy fish sizes and other size dependent research.

Two approaches were presented to estimate lens-subject distance to correct for parallax errors. Using calibration images was shown to be more accurate however this probably is not a general rule. The less accurate method calculated lens-subject distance from sensor size in real-world units and in pixels. The causes of error are likely incorrect manufacturers' figures for real world sensor size and focal length as the other quantities are known precisely. Empirical verification of which method best estimates lens-subject distance could be justified. Moreover, linear modelling of the error should produce a satisfactory correction.

5.5.1 Fiducial marker type

A foreground fiducial marker was shown to have the same accuracy as paired parallel lasers in the estimation of total length. Both markers are subject to the same underlying causes of error which arises when the mean RWLPP across the marker is not the same as the mean RWLPP across the dimension of the subject being measured. The background marker provided excellent estimates of total length, even in radially distorted images. The is because; (i) the RWLPP of the marker was calculated across the whole length of the subject and so mean RWLPP was the same irrespective of any distortions; (ii) when measuring fish length the caudal fin is at the same field depth as the background marker, hence there is only a small parallax error caused by the elevation of the snout above the background marker (explaining the overestimate). Using a background fiducial marker will only be accurate under specific conditions however, the general iterative approach to correct can equally be applied (with adjustments) when using a background marker. As with laser and foreground markers researchers must ask themselves what may cause the mean RWLPP to differ between the marker and the length being measured?

The choice of marker is context-dependent. Foreground and background markers are cheap but must be positioned close to the subject. Paired lasers can be projected onto a subject from a distance but become difficult to differentiate in strong sunlight or where the surface absorbs or diffuses the wavelength of incident laser light. Intense lasers also pose a health and safety risk. This renders lasers difficult to detect using machine vision. In the present study small fabrication errors in the laser housing and the laser unit itself were an additional cause of noise.

5.5.2 RADIAL DISTORTION

Lens distortion correction is available in multiple packages and the mathematics is well understood for tangential and radial distortion (Szeliski, 2010) however, lens distortion correction using OpenCV lens calibration achieved a reduction in RMSE to 0.76 pixels, which translates to submillimetre accuracy at lens-subject distances of between 192 mm and 659 mm. Neuswanger et al. (2016) reported improved correction performance with an extra parameter in the model. OpenCV is released under a 3-clause BSD license allowing the code to be modified, reused and redistributed hence it can be incorporated into machine vision pipelines, across diverse operating systems. Other methods of lens distortion correction tend to be in proprietary software released by the camera manufacturers (e.g. GoPro Studio), in photo editing software (e.g. Adobe Photoshop and PTLens), or in scientific software (e.g. MatLab). VidSync (Neuswanger et al., 2016) is an open source package which supports lens calibration however, it has no API and is authored in Objective C, hence can only be executed on Mac OS.

In the study by Hold *et al*. (2015), radial distortion was limited by ensuring the subject was centred in the field of view (also see Rogers *et al*., 2017). The parallax effect was empirically controlled with a 2nd order linear model with FoV as a predictor. The approach was successful in reducing bias however, empirical modelling is unsuitable where conditions cannot be prescribed. The model will only make known valid predictions over the quadvariate distribution of focal length, subject size, FoV (or lens-subject distance) and the subject rotation over which the model was fitted. Hence combining lens calibration with a mechanistic model of parallax effects (as presented), provides a generalizable solution which should be applicable to a wide range of length estimation correction when using fiducial markers.

5.5.3 Tangential distortion

Where the optical axis may not be perpendicular to the subject then tangential distortion must be corrected to minimise error. Chang *et al*. (2009, 2010) used a chessboard background-fiducial marker to estimate the TL of *Thunnus alalunga* from images with unknown tangential distortion. Tangential distortion was then corrected by manual review of each image and software was used to calculate the corrective affine transform. OpenCV provides support for calculating the corrective affine transformation and support for identifying chessboard vertices (or other regular structures). Using the OpenCV API would allow tangential distortion to be automatically corrected and the process can easily be incorporated into any image processing pipeline. The OpenCV ArUco marker library (Garrido-Jurado, Muñoz-Salinas, Madrid-Cuevas, & Marín-Jiménez, 2014) supports marker detection and predicts the affine transformation required to correct tangential distortion based on the orientation of the marker. However, applying the transformation will only reduce error if the subject plane and the marker plane are parallel, which may not always be the case.

5.6 Conclusion

The development of videogrammetry and photogrammetry undoubtedly has a role in meeting the increasing demands to gather ecological and fisheries data (e.g. European Commission, 2008). Additionally, the increasing availability and decreasing costs of robust cameras (Struthers et al., 2015) makes them more attractive to researchers (review Bicknell et al., 2016). This methodology shows the mechanistic corrections required to achieve accurate estimation of morphometric measurements from images captured with limited control over the equipment. Such correction would be a necessary step in automating the extraction of morphological data from images. Automating length estimation from images could reduce costs and would greatly increase the potential to collect finer grain data in population assessments, particularly—but not exclusively—in volunteer based projects. Automating image processing would also free time for more productive research activity.

Chapter 6

Using Machine Vision to Estimate Fish Length from Images

Chapter 6 is awaiting minor corrections in the journal *Methods in Ecology and Evolution* with some minor trimming of the text.

https://besjournals.onlinelibrary.wiley.com/journal/2041210x

KH, MK and FV critically revised the article. FV provided some methodological advice. All other work was that of GGM.

No highlights were required for this journal

6 Using Machine Vision to Estimate Fish Length from Images

6.1 ABSTRACT

A photograph encodes data which is frequently recorded in species observations. Examples include location, date-time, species, gender and morphometric measurements. Data on human activity are also encoded, such as the length distribution of fish removals from fishing activity. Much of the information is encoded as metadata and easily extracted however, morphological traits are not. Accurate length estimates from images can be made using a fiducial marker but the manual extraction of length estimates is time consuming and estimates can be inaccurate

when there is no control over the camera model used. This article presents a methodology which uses machine vision to estimate the total length (TL) of a fusiform fish (European sea bass). Three separate regional convolutional neural networks (R-CNN) were trained using public images of sea bass. Images of sea bass and a fiducial marker were captured using 3 non-specialist cameras. Images were undistorted using the intrinsic lens properties calculated for the camera, then TL was estimated using MV detection of the fiducial marker and the subject. Object detection and TL estimation bias were evaluated for the three R-CNNs under image downsampling and image rotation. Each R-CNN accurately predicted the location of fish in test images (mean intersection over union, 93%) and estimates of TL were accurate, with percent mean bias error (%MBE) [95% CIs] = 2.2% [2.0, 2.4]). Detections were robust to horizontal flipping and downsampling. TL estimates at absolute image rotations > 20° became increasingly inaccurate however, by using machine learning to remove outliers and model error %MBE [95% CIs] was improved to -0.1% [-0.2, 0.1] using the best overall performing R-CNN, NASNet (Zoph, B. & Le, Q. V., 2017. Neural Architecture Search […], in: Conference Proceedings ICLR). Machine vision can be used to classify and derive measurements of species from images. It is anticipated that ecological researchers and managers will be able to make increasing use of MV where image data is collected (e.g. in remote electronic monitoring, virtual observations, wildlife surveys and morphometrics) and will be of particular utility where large volumes of image data will be collected such as in national volunteer and citizen science programmes, data mining or long-term longitudinal studies.

6.2 INTRODUCTION

Currently only a small proportion of the world's marine stocks are sufficiently data rich to enable formal stock assessments to be performed hence most marine fisheries are data poor (Costello et al., 2012; Ricard et al., 2012). This is in spite of legislation which requires marine stocks to be exploited sustainably and managed with consideration of their associated ecosystems (e.g. United States Code, 1976, Stevens, 1996, European Commission, 2008). The potential for commercial fisheries to negatively impact stocks and ecosystems has long been known, but recreational fishing can also have a meaningful impact on fisheries and associated ecosystems (Cooke & Cowx, 2004; Lewin et al., 2006; McPhee et al., 2002). Marine recreational fisheries in particular can lack current and historical data even in developed countries (Hyder et al., 2018; ICES, 2017d). Furthermore many countries have had little or no catch recording, no registration or licensing schemes for some marine recreational fishing sectors, and no regular national directed surveys.

Fisheries assessments have survey phases in which a metrological measurement of the target species occurs (ICES, 2012; National Research Council, 2006; Pauly & Morgan, 1987). In both commercial and recreational fisheries, measurement has traditionally involved observations by researchers, fisheries managers or the fishers themselves. Observer costs are high in commercial monitoring (e.g. Needle et al., 2015) and in the assessment of recreational fisheries. Hence, there has been an increasing interest in remote electronic monitoring (REM, i.e. using images to record fish size, fish number and bycatch) (D. C. Bartholomew et al., 2018; Chang et al., 2010; Hold et al., 2015; Pasco et al., 2009; Rosen et al., 2013; White et al., 2006). Videogrammetry and photogrammetry (for brevity, simply photogrammetry) are also being used in non-destructive observational marine research (Deakos, 2010; Dunbrack, 2006; Harvey et al., 2001; Jaquet, 2006) and aquaculture (Costa et al., 2006; Tillett, McFarlane, & Lines, 2000; Zion, Shklyar, & Karplus, 1999; Zion et al., 2000).

The use of REM and related approaches is likely to increase as camera technology improves and equipment costs fall (reviews Struthers et al., 2015, Bicknell et al., 2016). Photogrammetry can reduce costs per data point (Chang et al., 2010) and could provide considerable savings when compared to observers (National Oceanic and Atmospheric Administration, 2015a). Capturing image frames produces vast volumes of data which is time consuming to process (e.g. Needle et al., 2015, van Helmond et al., 2017). This problem can be alleviated by using motion detection algorithm(s) to extract salient frames from videos (e.g. Weinstein, 2015), but extracted frames still requires manual processing to extract quantitative data. Object detection with machine vision (MV) could be used to automate the extraction of metrological data from images. Historically, MV has been used to analyse images which have been captured under controlled conditions (e.g. fixed cameras, backgrounds and lighting). This control makes the isolation of the subject from the background (segmentation) much easier, allowing computationally inexpensive techniques to be applied, e.g. using optical flow (Hsiao, Chen, Lin, & Lin, 2014; Spampinato, Giordano, Salvo, Fisher, & Nadarajan, 2010; Zion, Alchanatis, Ostrovsky, Barki, & Karplus, 2007) and segmentation by pixel properties (Jeong, Yang, Lee, Kang, & Lee, 2013; Lee, Schoenberger, Shiozawa, Xu, & Zhan, 2004; Strachan, 1993; White et al., 2006; Zion et al., 1999).

To date, photogrammetry has typically used multi-laser (D. C. Bartholomew et al., 2018; Bergeron, 2007; Deakos, 2010; Rogers et al., 2017) or multi-camera systems (Costa et al., 2006; Dunbrack, 2006; Harvey et al., 2001; Harvey & Shortis, 1995; Neuswanger et al., 2016; Rosen et al., 2013; Tillett et al., 2000), but the equipment is comparatively bulky and expensive. Single camera systems and a fiducial marker (i.e. an object of known scale placed in the image) have been used (Hold et al., 2015; Konovalov, Domingos, Bajema, White, & Jerry, 2017; van Helmond et al., 2017) but control of the camera model or the framing of the fiducial marker and subject was required (e.g. Muir et al., 2012, Rogers et al., 2017). Without this control, length estimates are subject to an unknown error because of the intrinsic lens differences between cameras (e.g. radial distortion and parallax effects). The additional challenges in extracting quantitative data from images taken by volunteers (or where expensive or less portable equipment is unsuitable) may explain the almost complete lack of a suitable solution. Over the last 6 years, convolutional neural networks (CNN) have become the top performers in object detection. CNNs are dominating competitions such as ImageNet LSVRC challenge (IMAGENET, 2018) and displacing feature based methods such as SIFT (Lowe, 2004) and HOG (Dalal & Triggs, 2005). The success of CNNs has stimulated efforts to develop open-source application programming interfaces (API) and the techniques are now mature and stable enough to be viable solutions for competent programmers to perform object detection using regional CNNs (R-CNN).

This article aims to explore the feasibility of using MV to automate the identification and size estimation of an important species from images. The objectives are to (i) introduce the software and methods to achieve length estimation with a cheap and portable fiducial marker; (ii) to show that length estimates can be made with no control over image background, lighting or specialist cameras using a foreground fiducial marker; (iii) provide region of interest labelled images of the European sea bass, *Dicentrarchus labrax* (see Appendix K); (iv) to compare the speed and performance of multiple R-CNN networks.

6.3 METHODS[4]

6.3.1 ETHICS

Sea bass captures were made by recreational fishers and a commercial vessel as part of their day-to-day activity. All reasonable measures were taken to minimise air exposure time to the fish while photographs were taken. Ethical approval was granted by the Animal Welfare and Ethical Review Board of Bangor University, Wales, UK.

6.3.2 TRAINING AND VALIDATION IMAGE ACQUISITION

Training (n = 734) and validation (n = 184) images were obtained from public sources and the region of interest (RoI) for each image was drawn tight to the sea bass body, to the limits of the caudal fin tip(s) and the snout vertex (Figure 6-1). Training and inference was carried out in Tensorflow (Google, 2018) using transfer learning from 3 pretrained R-CNNs. The networks were (i) ResNet-101 (He, Zhang, Ren, & Sun, 2016), (ii) Single shot MobileNet detector (A. G. Howard et al., 2017) and (iii) NASNet (Zoph & Le, 2017); abbrevs. ResNet, MobileNet and NASNet respectively. Intersection over Union (IoU) is reported as an indicator of object localization accuracy. Each model outputs an objectness score (*score*) which is interpreted as the probability that the MV proposed region contains the predicted class (Ren, He, Girshick, & Sun, 2017).

6.3.3 FIDUCIAL MARKER SELECTION AND IMAGE ACQUISITION

Three ArUco fiducial markers (Garrido-Jurado et al., 2014) of side lengths 25mm, 30mm and 50mm were mounted on polypropylene sheets (Figure 6-2). Photographs of sea bass were taken with the informed consent of fishers on the shore and afloat using 3 different non-specialist cameras (henceforth *marker images*) at fixed focal length. Fish were posed to minimise body distortion and occlusion. Fish total length (TL) was measured and recorded.

Figure 6-1. Defined region of interest (RoI) for training and validation images.

Top, both tips of the caudal fin included in the RoI set tight to the lower and upper bounds of the body depth. Bottom, single caudal fin tip included. Images where no caudal fin tip would fall within the tight body bounding box were rejected.

[4] Appendix J contains additional methodological detail.

The marker was positioned on the fish as in Figure 6-3 and then photographed.

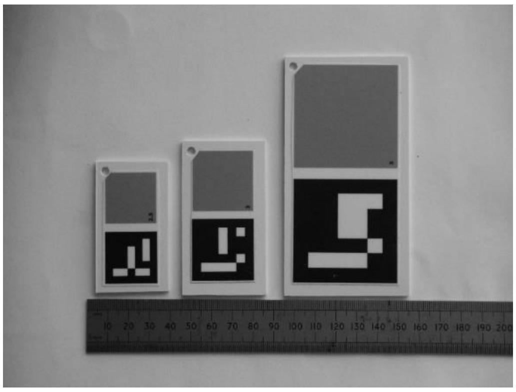

Figure 6-2. ArUco fiducial markers (Garrido-Jurado et al., 2014) mounted on polycarbonate. Left to right, 25 mm, 30 mm, 50 mm. The unique marker pattern is identified as part of the detection, hence the world marker length can also be determined.

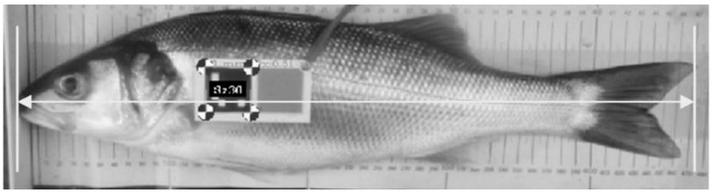

Figure 6-3. Placement of ArUco marker (Garrido-Jurado et al., 2014) during image capture.

6.3.4 UNDISTORTING MARKER IMAGES

Images from each of the 3 cameras were corrected for radial and tangential distortion with the OpenCV API (OpenCV team, 2018; Scaramuzza, Martinelli, & Siegwart, 2006; Z. Zhang, 2000). Lens calibration profiles were created in OpenCV for each camera at each supported field of view and mechanical zoom (henceforth *undistorted images*).

6.3.5 LENGTH ESTIMATION

An R-CNN attempts to predict the rectangle which most accurately bounds the subject within the image and outputs the detection rectangle defined by its 4 vertices. When estimating TL, the pixel length of the long side of the detection rectangle approximates to the TL (pixels) of the sea bass. The world length per pixel, $\bar{l}$ was estimated from the 4 sides of the detected ArUco marker according to, $\bar{l} = \frac{1}{n} \cdot \sum_{1}^{n} l/p_i$ where p_i is the i^{th} side length in pixels, and l is the real side length (e.g. 50mm). The accuracy $\bar{l}$ was validated manually (Linear Regression, b=1.003, R^2=0.999) using ImageJ (Schneider et al., 2012). Mean absolute error (MAE) and mean bias error (MBE) are reported, they are calculated as follows, $MAE = \frac{1}{n} \cdot \sum_{i=1}^{n} |l_i - \widehat{l_i}|$ and $MBE = \frac{1}{n} \cdot \sum_{i=1}^{n} l_i - \widehat{l_i}$ where l_i is the i^{th} estimate of TL and $\hat{y}_i$ is the expected (i.e. actual) TL of the i^{th} element. Hence a negative bias would represents an underestimate of TL.

6.3.6 Detection and length estimation under rotation, flipping and downsampling

The accuracy of TL estimates under three translations were checked, these were; (i) image rotation between -30° and 30° in increments of 1°; (ii) horizontal flipping around the line $x = 0.5 \cdot width$; and (iii) downsampling by a factor 1.5 to a minimum image height or width of 50 pixels. TL estimates for rotated images were corrected based on the geometry of the detection box under increasing rotation in relation to the snout and caudal vertices of the subject.

6.3.7 Removing outliers and modelling bias

NASNet R-CNN detections were split into train and test data. Training data were used to identify biased outliers using an isolation forest (Liu, Ting, & Zhou, 2008; Pedregosa et al., 2011) with the variables, (i) ratio of height to width of the detection, (ii) objectness score and (iii) %MBE. Outliers were then removed from the training set, and a gradient boost regressor (Friedman, 2002; Pedregosa et al., 2011) trained predictors on (i) and (ii) above. Outliers were removed from the test dataset and the gradient boost model used to correct bias. Further methodological details are given in Appendix L.

6.3.8 Reported length estimates

Several length measurements are reported, roughly in increasing order of complexity. The methods for calculating *corrected MV-TL* are those given for variable *profile corrected TL* in 5.3.4 and 5.3.5. All TL estimates are based on undistorted images (except *physical-TL*):

(i) *Physical-TL*. The direct measurement of the physical fish with a ruler.

(ii) *Corrected manual-TL*. Manual estimation of the marker and fish length from the undistorted image with ImageJ. Parallax corrections applied (Appendix J).

(iii) *MV-TL*. MV estimates of TL on undistorted images with no other corrections.

(iv) *Corrected MV-TL*. MV-TL plus a correction for parallax errors (Appendix J).

(v) *Rotation corrected MV-TL*. Corrected MV-TL plus a geometric correction based on the height and width of the detected region (Appendix J, 1.4.3) to adjust for detections under rotation.

(vi) *Model corrected MV-TL*. Rotation corrected MV-TL plus correction with machine learnt models generated from training data to remove outliers and correct bias in test data (Appendix J, 1.6). Only test data reported.

Code and images with the ground truth rectangles defined in the VGG Image Annotator (http://www.robots.ox.ac.uk/~vgg/software/via) are published at https://github.com/seabass-detection/seabass-detection.

6.4 RESULTS

For every non-transformed sea bass image, each CNN generated region proposals with objectness scores > 0.5 (with the exception of a single MobileNet score of 0.01). All regional proposals were at least partially coincident with ground truth, with a minimum IoU of 45% (Figure 6-4). Negative images had no false detections under any network, (score mean ±S.D. of 0.005 ±0.008, n = 30, max = 0.04). N.B. all remaining means report S.D. unless otherwise specified.

Figure 6-4. Minimum observed intersection over union (IoU=45%) in a marker image. Machine vision detection in green.

Detection performance between networks was practically indistinguishable for inference on untransformed images (mean IoU; NASNet 93.5% ±2.5; ResNet, 92.5% ±6.2; MobileNet, 92.2% ±3.5) and horizontally flipped images (NASNet 93.3% ±2.2; ResNet, 93.4% ±5.1; MobileNet, 92.8% ±3.0) hence it is reasonable to assert that detections were effectively invariant to horizontal flipping (%IoU mean [95% CIs]; horizontal flip, 93.2% [93.0, 93.4]; untransformed, 92.8% [92.5, 93.0]). This equivalence is despite the large differences in mean detection times per 1000 pixels^2 (NASNet 11.4s ±0.005; ResNet 4.1s ±0.608; MobileNet, 1.1s ±0.034). However, when visualised it is apparent that the NASNet network delivered more consistent object detection with no IoU outliers (Figure 6-5). All single shot detector detections had IoUs > 75%, however ResNet had 7 detections < 75% IoU (1.1% of all detections).

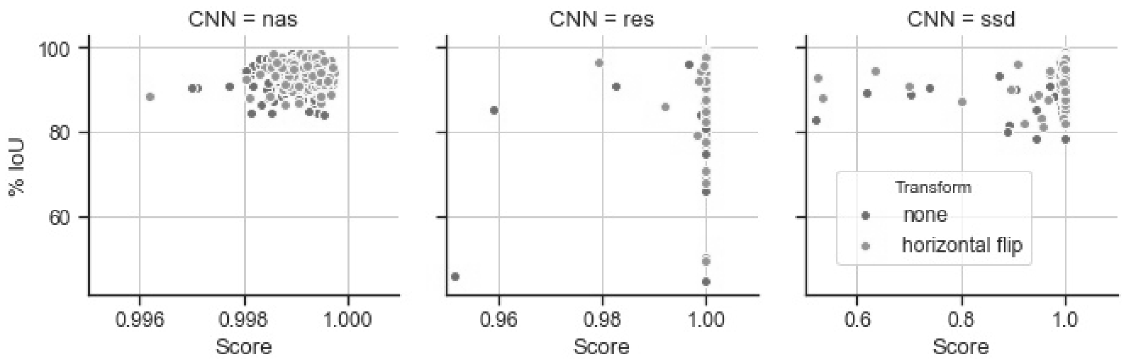

Figure 6-5. Objectness score *vs*. intersection over union for untransformed and horizontally flipped images under detection from NASNet (Zoph & Le, 2017), ResNet-101 (He et al., 2016) and single shot MobileNet detector (A. G. Howard et al., 2017) regional convolutional neural networks (Google, 2018). Note that the x-axis scales increase by approximately two orders of magnitude from left to right. Objectiveness scores < 0.5 were excluded.

6.4.1 Length estimates

Figure 6-6. Two images in which the ArUco marker (Garrido-Jurado et al., 2014) was undetectable.

ArUco markers were consistently recognised using the OpenCV API under natural conditions, with the marker successfully localized in 99.3% of untransformed images. Two detection failures occurred because of over-exposure (Figure 6-6). *Corrected MV-TL* estimates had a MBE of 5.9 mm ±20, compared with MBE derived from *corrected manual-TL* estimation of -0.5 mm ±14.8. C*orrected MV-TL* estimates showed consistent variance in bias across *physical TL* (Figure 6-7). On excluding TL estimates made under the noisier ResNet and MobileNet networks, MBE for *corrected MV-TL* estimates was increased by 2 mm to 7.9 mm nevertheless, S.D. decreased to 14.7 mm, matching the precision of manual estimates of TL (*corrected manual-TL*).

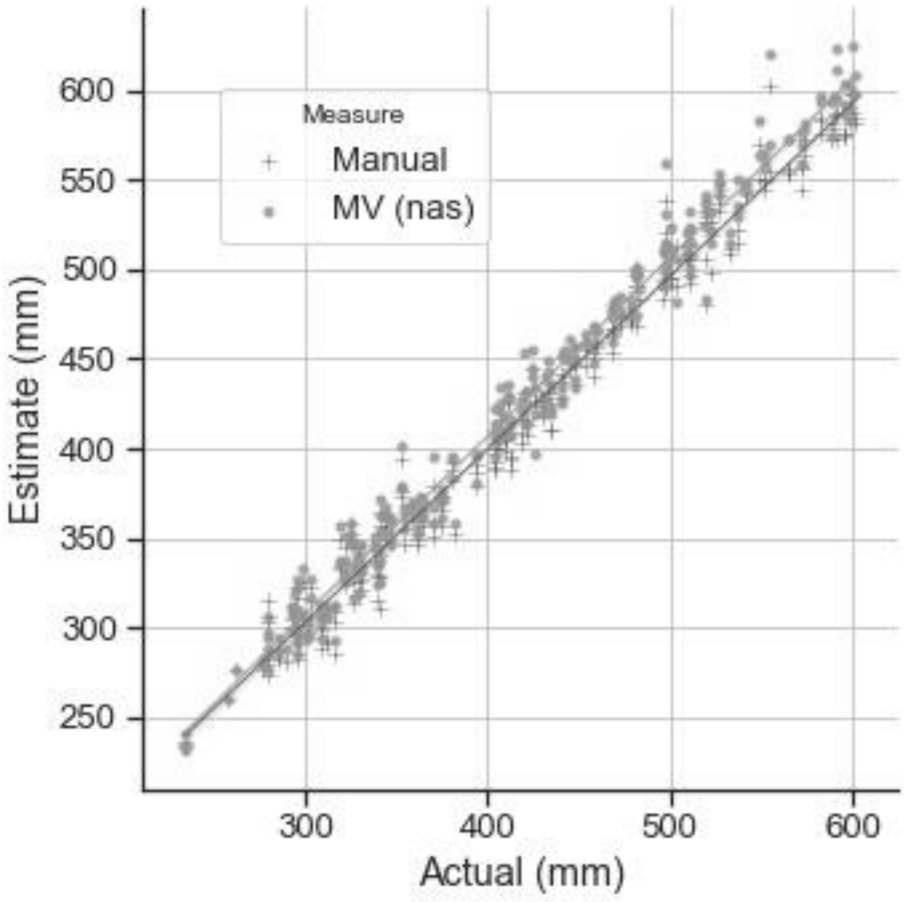

Figure 6-7. *Physical total length* of sea bass *vs*. estimated length using manual measurement (Manual, *corrected manual TL*) and machine vision (MV nas, *corrected MV-TL*), NASNet CNN only (Zoph & Le, 2017).

Manual and MV TL estimation errors tended to be less accurate and precise (mean squared error, MSE) when made on the shore rather than afloat (Figure 6-8, MSE; Afloat, 7.9; Shore, 25.9), and there was no apparent systematic bias in length estimation introduced by the camera model when comparing *corrected manual-TL* estimates (which have lower variance than MV length estimates) with platform as a covariate (ANCOVA, $F_{(2, 1787)}$, $p = 0.15$). Mean %MBE for manual TL estimates were 0.7% ±4.6, 1.1% ±4.0 and 0.7% ±4.1 for the GoPro, s5690 and XP30 cameras respectively.

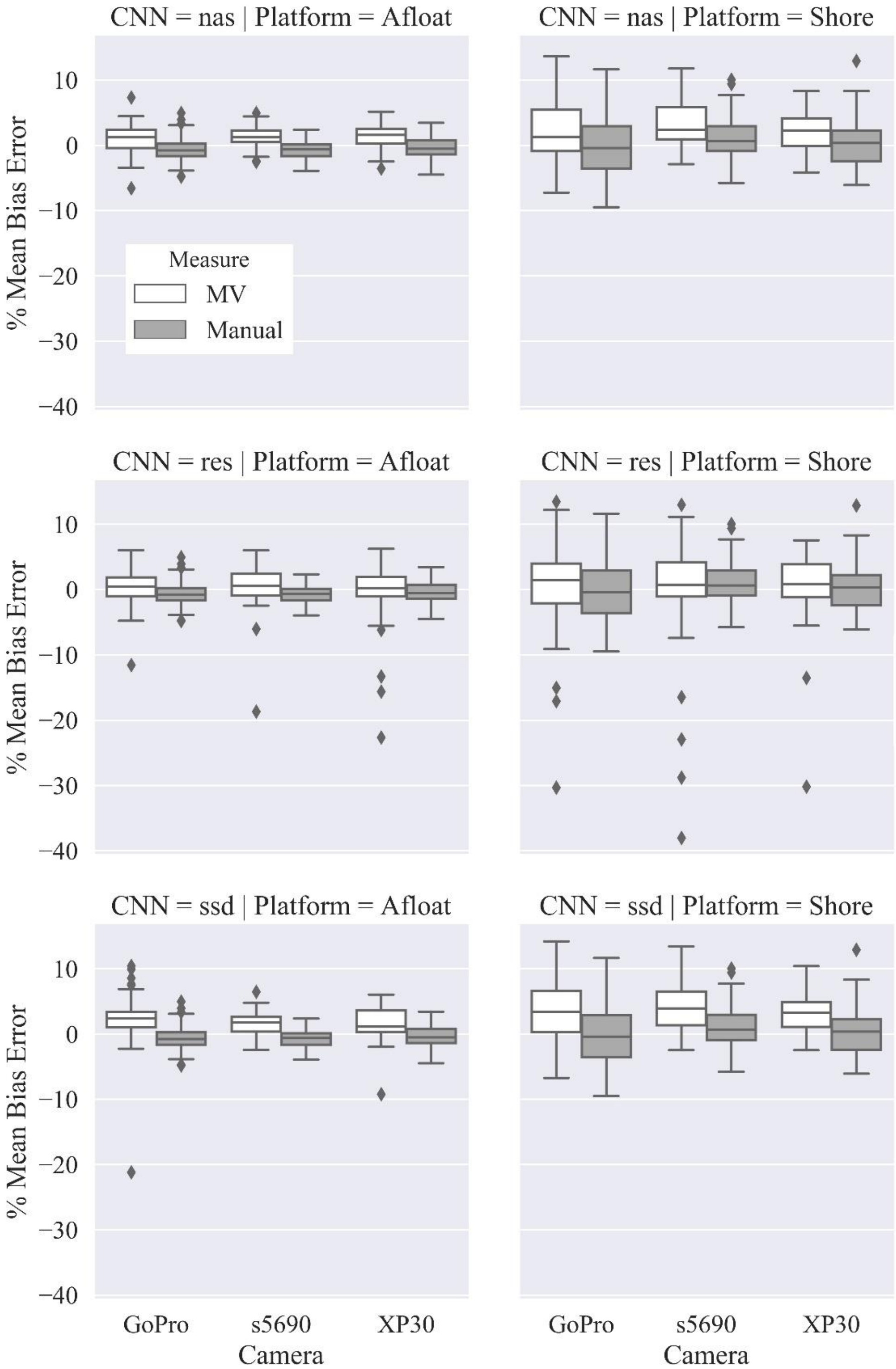

Figure 6-8. Mean bias error of sea bass total length estimates made manually (*corrected manual TL*) and using machine vision (*corrected MV-TL*). Machine vision estimates were made using 3 different regional convolutional neural networks, NASNet (Zoph & Le, 2017), ResNet-101 (He et al., 2016) and single shot MobileNet detector (A. G. Howard et al., 2017). Images were captured onboard boats (Afloat) and from the shore using 3 different cameras.

The increased %IoU outliers observed during detection with ResNet and—to a lesser degree—the MobileNet single shot detector—manifest as the %MBE outliers in Figure 6-8. The ResNet detector produced 9 of the top 10 MV associated underestimates (fully corrected percent errors -16.4% to -38.0%). These errors arose because detections followed the approximate pattern observed in Figure 6-4, with the ResNet detector occasionally truncating the detection. This behaviour was not observed in the other detectors on untransformed images (i.e. an image which has not been flipped, downsampled or rotated).

6.4.2 SCALE

ArUco marker detection was robust to downsampling to approximately 30% of the original image size (original image size, mean = 1355 by 1029 pixels, or 1.5M pixels^2). ArUco markers were approximately 18 pixels^2 at 30% of original image size and images were approximately 400 by 300 pixels (120k pixels^2). At 30% image size the marker detection rate was 93% however, this dropped to 53% at the next scaling factor of 20% (Table 6-1).

Table 6-1. ArUco fiducial marker (Garrido-Jurado et al., 2014) detection rates under image scaling (factor = 1.5) with width and height minimum limit of 50 pixels.

Marker size is the average side length of the marker in the image. G.T. width is the ground truth horizontal length. Columns are means ±S.D. Obj. score is the mean objectness score across all networks. ND = no detections, px = pixels.

Scale	N	Width (px)	Height (px)	Marker size (px)	G.T. width (px)	Obj. score	% Det.
1	921	1,355	1,029	63 ±15	874 ±132	1.00 ±0.04	100.0
0.67	921	903	685	42 ±10	536 ±79	1.00 ±0.02	99.3
0.44	921	601	456	28 ±6	357 ±53	1.00 ±0.04	98.7
0.30	921	400	303	18 ±4	238 ±35	0.99 ±0.04	92.8
0.20	921	266	201	13 ±3	158 ±23	0.98 ±0.10	52.8
0.13	921	177	133	10 ±3	105 ±15	0.91 ±0.21	13.0
0.09	921	118	88	7 ±1	70 ±10	0.77 ±0.34	1.3
0.06	918	78	58	ND	47 ±7	0.55 ±0.39	ND
0.04	3	62	50	ND	26 ±0	0.005 ±0.007	ND

The networks on average maintained objectiveness scores of ~98% at the 20% scaling factor, where the mean image size was 41.4k pixels^2 (i.e. ~203 pixels^2). At this image size, the average ground truth RoI was 158 by 23 pixels (Table 6-1). NASNet produced marginally more accurate TL estimates under downsampling. For each network %MAE increased in increments of between 1% and 2% until the downsampling factor exceeded ~30% (mean ground truth width = 238 pixels), after which %MAE began to increase in larger increments. Each network responded similarly to downsampling (Figure 6-9), at 20% image size, %MAE = 9.9% ±7.8 which increased markedly to 15.9% ±8.4 at 13% of original image size at ~153 pixels^2.

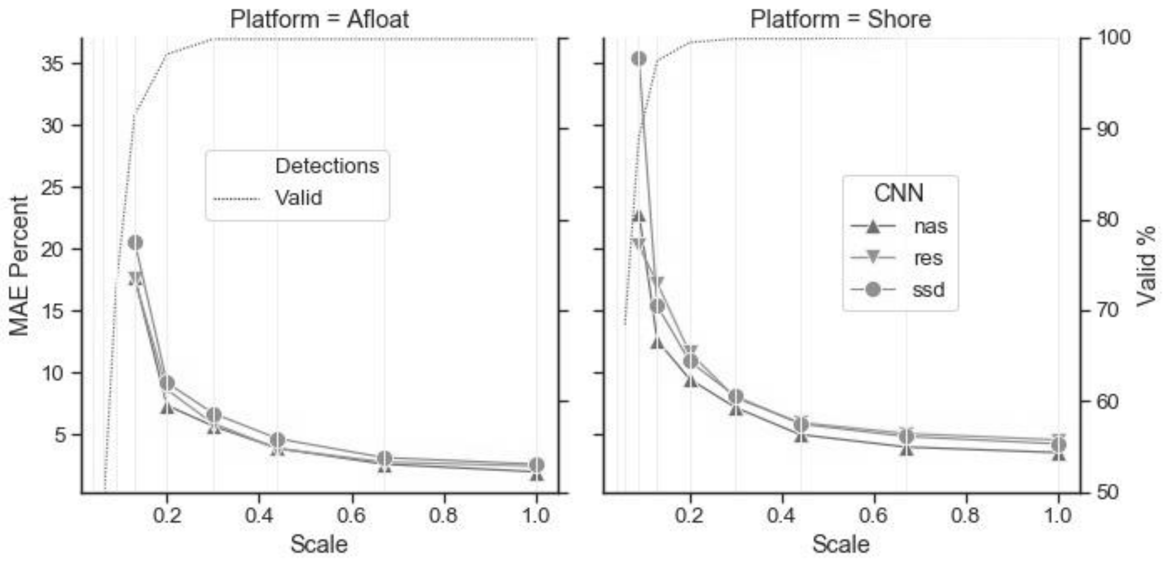

Figure 6-9. Effect of downsampling images by factor *Scale* on the percent mean absolute error (MAE) of length estimates (*corrected MV-TL*) for the 3 regional convolutional neural networks, NASNet (Zoph & Le, 2017), ResNet-101 (He et al., 2016) and single shot MobileNet detector (A. G. Howard et al., 2017). MAE excludes length estimates below objectness scores of 0.5, = percentage of detections with objectness score > 0.5. Grey horizontal lines indicate scale factors.

6.4.3 Rotation

The NASNet and ResNet networks behaved similarly under image rotation (Figure 6-10) and detection was robust to small amounts of rotation, with over 90% of objectiveness scores greater than 50% at absolute rotation ≤ 20% for the NASNet and ResNet networks. At 20° absolute rotation the MobileNet network had 67% of objectiveness scores below 50%. As the absolute rotation angle increased beyond ~15°, NASNet and ResNet predictions of *corrected MV-TL* exceeded 5% %MBE however, %MBE was 2.5% for the MobileNet network (Figure 6-10, absolute rotation = 15°, %MBE [95% CIs]; NASNet, -5.0% [-5.3, -4.6]; ResNet, -5.3% [-5.9, -4.7]; MobileNet, 2.7% [2.2, 3.3]). This apparently good performance of the MobileNet CNN masks the greatly decreased confidence in regional proposals under the MobileNet network (score series, Figure 6-10) and a corresponding loss of valid detections.

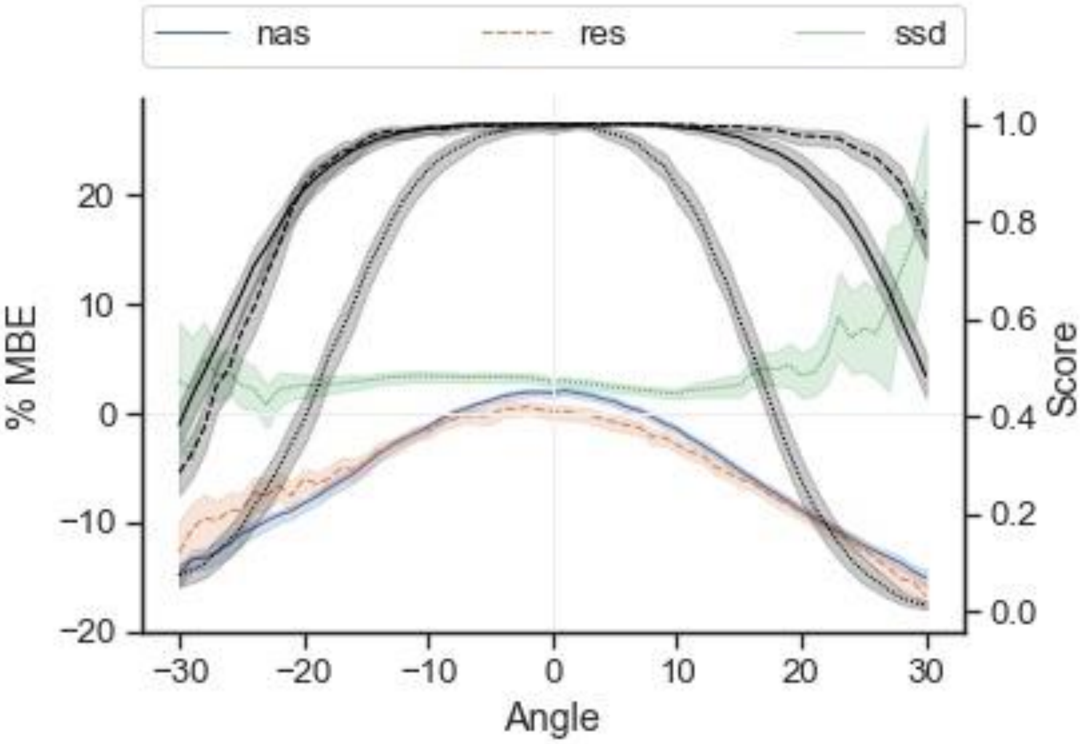

Figure 6-10. Effect of rotation on percent mean absolute error (% MBE) with 95% confidence intervals for machine vision derived estimates of total length of European sea bass using the following pretrained regional convolutional neural networks, NASNet (Zoph & Le, 2017), ResNet-101 (He et al., 2016) and single shot MobileNet detector (A. G. Howard et al., 2017). % MBE excluded objectness scores < 0.5. Black lines are all objectness scores for each network and share the corresponding line style i.e. =ResNet (ResNet), =MobileNet (MobileNet), =NASNet (NASNet). Note this shows the variable *rotation corrected MV-TL* which performed marginally better than *corrected MV-TL* with decreased %MBE for NASNet and ResNet TL estimates.

The geometric rotation correction (variable *rotation corrected MV-TL*) did not consistently decrease bias for all rotations (see supplementary data I) and bias reduction was only marginally improved for the NASNet and ResNet networks (by 1.2% and 0.5% respectively) but bias was increased for the MobileNet network (1.0%). The failure to reduce bias to a constant across the ±30° rotation range is attributable to the divergence of the geometric model (detailed in Appendix J) from the bounding features of the subject which the CNNs “chose” under rotation. In essence, the CNN detections cannot be represented by the simple geometry of a rotating rectangle (Figure 6-11). Both the NASNet and ResNet networks displayed a consistent hyperbolic pattern in TL estimation bias through the rotation range and consistent variance in predictions across this range, whereas the MobileNet network did not share this consistency (Figure 6-10).

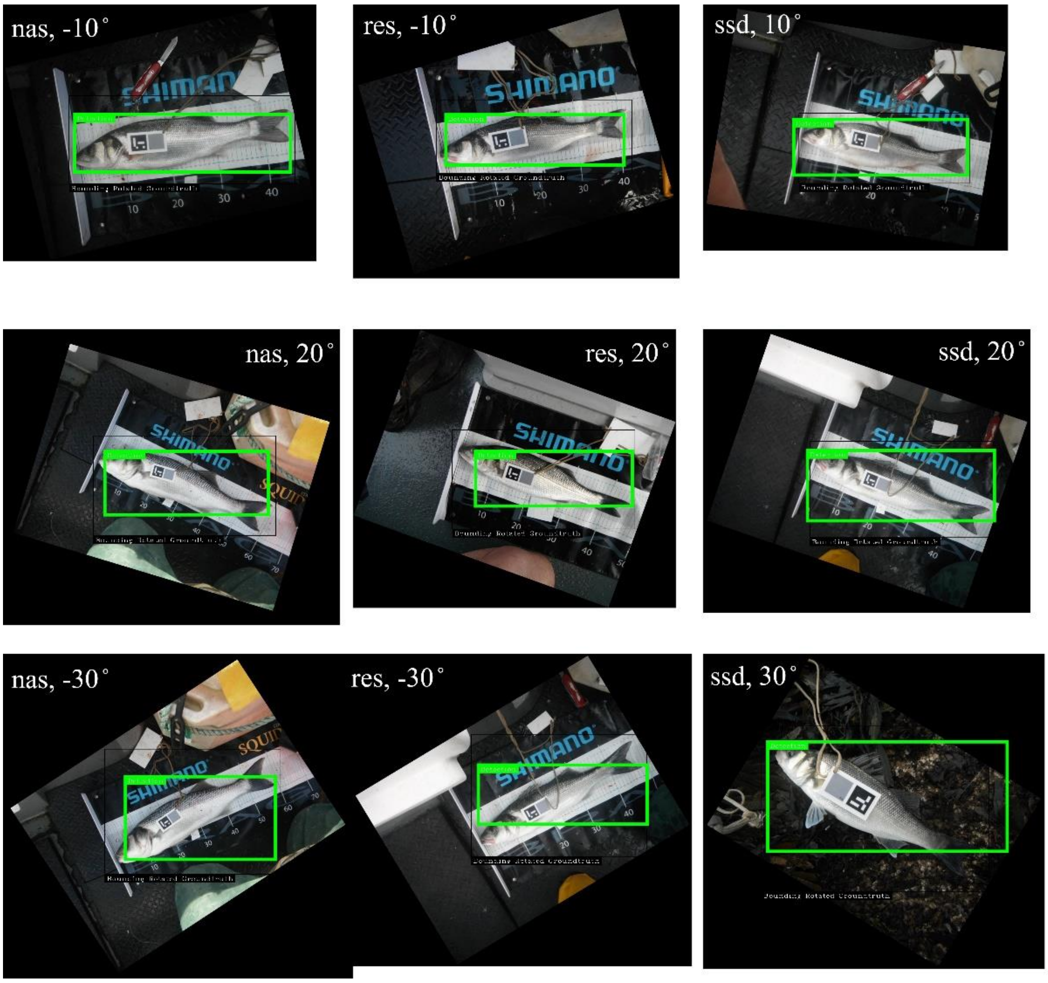

Figure 6-11. Example detections of sea bass under rotation for NASNet (Zoph & Le, 2017) , ResNet (He et al., 2016) and MobileNet (A. G. Howard et al., 2017) regional convolutional neural networks. Machine vision detections are visible in green.

Combining outlier removal and adjusting *rotation corrected MV-TL* per sample using the trained gradient descent regressor model produced a marked reduction in %MBE across rotations (Figure 6-12; Mean [95% CIs]; Corrected, -4.1% [-4.3, -3.9]; Model corrected, -0.5% [-0.6, -0.3]). Model correction also significantly centred bias at ~0% for absolute rotations $\leq 20^{\circ}$ (Mean [95% CIs]; *rotation corrected MV-TL*, -2.0% [-2.2, -1.9]; *model corrected MV-TL*, -0.1% [-0.2, 0.1]). The overall improvement on applying all corrections to MV estimates following lens correction only are unambiguous, with *MV-TL* estimates of %MBE [95% CIs] = -11.4% [-11.6, -11.2].

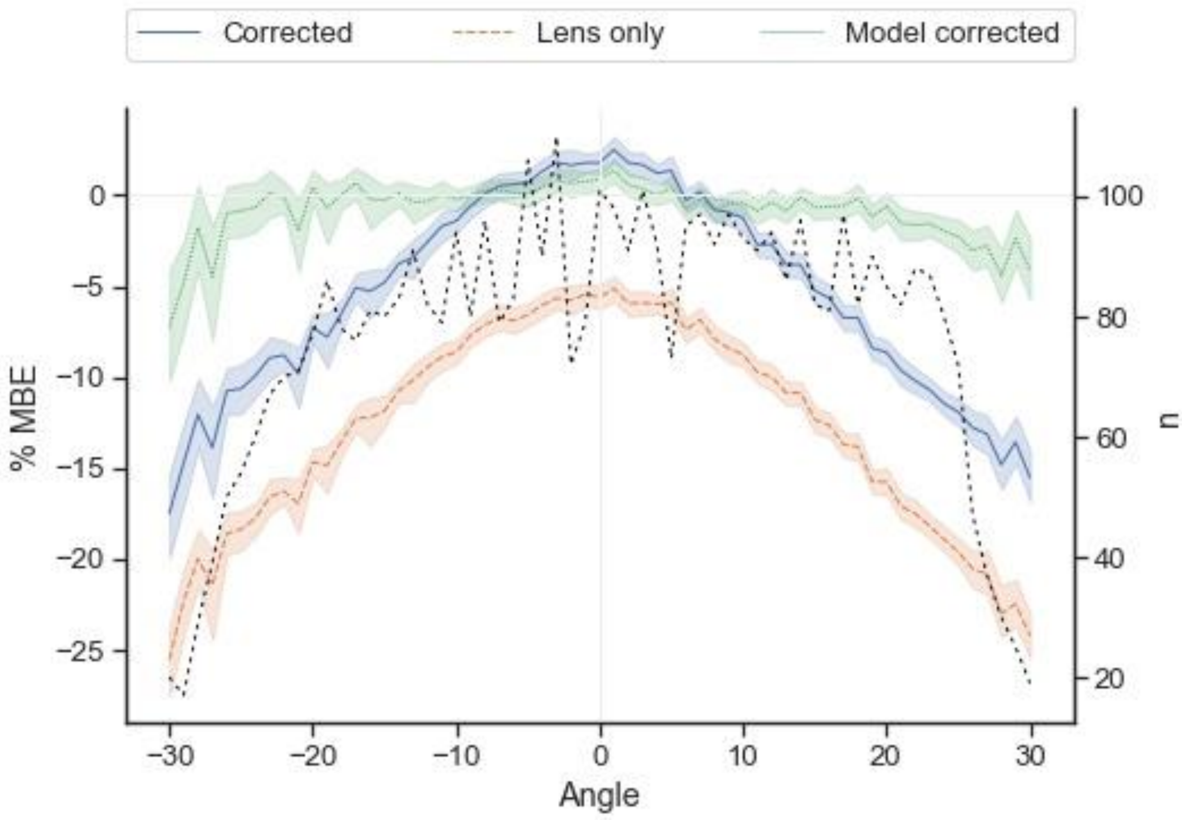

Figure 6-12. Progressive reductions in % mean absolute error (% MBE) for machine vision length estimates under experimental image rotation on test data using correction and outlier modelled on training data (partitioned at 7:3 ratio). Corrections are (i) camera lens correction only (, i.e. variable *MV-TL*); (ii) corrected ( ), lens correction and correction for the discrepancy between the proximal plane and the change in subject depth across the measured dimension of the subject (i.e. variable corrected *MV-TL*) and (iii) model corrected ( ), all corrections with additional removal of outliers identified with an isolation forest and bias prediction with a trained stochastic gradient descent regressor model (i.e. variable *model corrected MV-TL*). Series is the absolute number of samples (n) at each rotation. 95% confidence limits are shaded.

6.5 Discussion

This study introduced an end-to-end methodology to estimate total fish length—a crucial measurement in multiple stock assessment methods (Pauly & Morgan, 1987)—using recent advances in open-source R-CNNs and associated software applications (e.g. Abadi et al., 2015, OpenCV, 2018). By using these resources, it was shown that the position of an organism in an image could be accurately predicted without strict control over lighting conditions or subject background, both of which are usually controlled to simplify machine vision segmentation of fish species (e.g. Nery et al., 2005, White et al., 2006, Zion et al., 2007, Chuang et al., 2011, Jeong et al., 2013, Miranda and Romero, 2017). The high degree of accuracy of the predicted RoI (> 90% IoU) enabled the accurate estimation of an important morphological trait. Length estimation was achieved without reliance on specialist cameras, which are typically used in photogrammetry and videogrammetry to minimise radial distortion. Length estimates were also made without deploying multi-camera systems (e.g. Harvey et al., 2001, Costa et al., 2006, Dunbrack, 2006, Rosen et al., 2013) or paired lasers (e.g. Deakos, 2010, Rogers et al., 2017). Captured images are also provided for the use of others to train MV models to identify European sea bass, an important commercial and recreational marine species.

6.5.1 Networks

Of the three networks NASNet outperformed the ResNet-101 and MobileNet networks and was particularly effective at limiting outlier detections. However, the NASNet network had the slowest detection speeds of the three and was the most resource intensive. To execute transfer learning, batch size for NASNet had to be limited to 1 to fit within the 6 Gb of memory of the NVIDIA 1060 GTX card (configuration files are available at https://github.com/seabass-detection/seabass-detection/tree/master/tf_config). This is unsurprising as the NASNet has many more parameters than ResNet (Zoph & Le, 2017).

Neither ResNet nor NASNet are currently capable of performing real-time detections however, MobileNet can be deployed on mobile devices. The performance of MobileNet in this task was arguably better than ResNet and real time detection would be of particular benefit in volunteer based data collection applications where users could be given immediate feedback on species identification (e.g. Kumar et al., 2012, Fishbrain, 2018, International Game Fish Association, 2018, Stowell, 2018) and the success or failure of a particular recognition task.

6.5.2 LENGTH ESTIMATION

There is evidence that using cameras instead of active observers could reduce biases with the presence of observers affecting fisher behaviour (Benoît & Allard, 2009; Faunce & Barbeaux, 2011). Automating length estimates can mitigate self-sampling biases and also reduce the perception of researchers that measurements reported by non-scientists are biased or not recorded as rigorously as by trained observers or researchers (Kraan et al., 2013). Any digit biases, a persistent problem in human measurement (Tarrant & Manfredo, 1993) will be negated. Increased sample throughput could reduce biases which can arise through subsampling processes (e.g. Kraan et al., 2013). Images also have the benefit of providing an archive of data. However, problems can occur when deploying virtual observations and images or movies can suffer from poor quality, obscuring of the lens, or equipment failure, which can result in missing observations (e.g. Needle et al., 2015, van Helmond et al., 2017).

Length was more accurately estimated on afloat platforms than on the shore, this is because the afloat platforms provide a flat surface on which to measure and photograph the subject. Across both platforms and all camera models there was a small but consistent overestimate of size (mean bias error, 1.6%; 6 mm). Possible explanations include an underestimate of lens-subject distance during the camera calibration process which did not account for the internal distance between the lens and the glass cover of the cameras, or incorrect estimation of the parameters (e.g. mean profile height) used in the length correction calculation.

Bias magnitude was consistent across the range of fish lengths measured (~25 cm to 65 cm) hence a small correction could be calculated empirically during model training. The model used for rotation correction was successful in practically eliminating bias (% MBE = -0.1%), which brought the error magnitude in line with methods which control the imaging conditions (Hold *et al*. 2015, 0.1% in *Cancer pagurus, 0.6%* in *Homarus gammarus*; White *et al*. 2006, 0.3%, in *Hippoglossus hippoclossus*), use paired lasers (Deakos 2010, 0.4% in *Manta alfredi*) or multiple cameras (Harvey *et al*. 2001 -8.6% in artificial fish proxies; Rosen *et al*. 2013 1.0% across *Scomber scombrus*, *Pollachius virens* and *Pollachius pollachius*). This degree of bias is within the most stringent confidence level (Level 3, ±5% C.I.) defined for EU fisheries data collection (European Commission, 2013, 2016a).

Despite bias being largely eliminated, outliers in TL estimates were observed (minimised under NASNet). Without rotation, this error was largely attributable to errors arising from the subject pose in the image. Parallax errors arising through depth differences across the fiducial marker and the subject will be a major source of error which are typically dealt with by excluding images following manual review (e.g. Jaquet, 2006, Deakos, 2010, Rohner et al., 2015, Rogers et al., 2017). Correction for tangential deflection of MV designed fiducial markers is generally supported (e.g. Bergamasco et al., 2011, Garrido-Jurado et al., 2014), but this is unlikely to be a consistent correction (pers. observ.) for foreground fiducial markers where the tangential displacement of the marker differs from that of the subject.

Another source of error arises where there is curvature of the body of the subject. Length estimation of non-linear body pose can be made by identifying depth midpoints and calculating the line bisecting these midpoints (Strachan, 1993; White et al., 2006) or line fitting to subject contours (Miranda & Romero, 2017). Both approaches require accurate segmentation of the subject from the background, which is problematic in uncontrolled environments. Nevertheless, Tensorflow provides pretrained R-CNNs capable of segmentation (Google, 2018; He, Gkioxari, Dollar, & Girshick, 2017) and real-time segmentation is possible (e.g. Paszke et al., 2016). Advances in segmentation are a promising area for improving length estimates under body curvature as commonly observed in live shark specimens, remote electronic monitoring and terrestrial species. Alternative approaches may include training the network to identify sub regions of the subject which are subject to less distortion (e.g. the head) and keypoint detection (Kazemi & Sullivan, 2014; Y. Sun, Wang, & Tang, 2013; Vandaele et al., 2018).

6.5.3 Transformations

Detections and length estimations were robust to flipping and downsampling. Under decreasing image size the fiducial marker was found to be the limiting factor for the automatic extraction of TL. This is an intrinsic limitation of using a foreground fiducial marker where increasing marker size could obscure salient features. The lowest IoU was observed on the smallest sea bass sampled, where the marker occluded a comparatively large proportion of the subject (Figure 6-4). The effectiveness of the CNN under substantial downsampling does indicate that image sizes can be significantly reduced prior to inference to improve speed and reduce memory requirements.

Length estimates were unbiased and acceptably precise at small degrees of rotation. The bounding box under rotation generally predicted the x-coordinates of the snout and caudal vertices reasonably well, particularly under the NASNet network (Figure 6-11). Unfortunately the simplistic geometric model used (Appendix J, 1.4.3) largely failed to adequately correct length estimates under rotation. This failure to demonstrate generalizability through all rotations poses a serious limitation in some deployment scenarios. Under volunteer targeted image collection, a significant proportion of subject rotations would exceed the experimental rotation limits. A common and trivially implemented (but computationally expensive) approach to achieving rotation invariance is the brute force repetition of detection through incremental rotations with the best detection determined by some metric (or combination of metrics), such as the objectness score, detection height to width maxima or maximal agreement with a secondary detection technique such as template matching (Brunelli, 2009). In the present study accurate detections were achieved at absolute rotations to ~15° which suggests that steps of 15° could be used to reduce the search space. However, it may prove to be more efficient to train the network on incrementally rotated images, something again which is relatively trivial and has native support in most CNN APIs. However, data on rotation invariance under rotated training images was not published by Zoph and Le, (2017) and R-CNNs are not intrinsically rotation invariant, hence further empirical investigation is required

6.5.4 Other applications

The fiducial marker deployed was particularly easy to identify in fully automated MV processing pipelines and performed well as evidenced by the low bias and high detection rates. The choice of a foreground marker was driven by the use case, i.e. in large scale volunteer based surveys and data gathering exercises. A foreground marker is cheap and portable, and volunteers cannot deliberately inflate size estimates by moving the marker further away from the subject as would occur with a marker on which the subject is placed. The principles of the method are applicable to any type of marker (which can be detected) and multicamera systems, and to any organism for which morphological estimates are made, provided sufficient volumes of data are to be collected to justify technical development. Difficulties will arise in unconstrained camera systems where the scale indicator is difficult to distinguish in the image, such as lasers in strong sunlight (pers. observ.). Non-specialist markers can be segmented and length estimated using machine vision, such as a standard ruler (Konovalov et al., 2017). Opportunistic fiducial markers could also be segmented (e.g. human face) and used to produce estimates of fish size from historical images as has been done manually to provide ecological data on some species (Belhabib et al., 2016; Canese & Bava, 2015; McClenachan, 2009; Rizgalla et al., 2017).

The recent advances in the accuracy of MV mean that commercial length estimates may be made without the need for complex and costly mechanical pre-sorting and identification may be possible under occlusion by combining R-CNNs and deformable parts models and landmark detection (Felzenszwalb, Girshick, McAllester, & Ramanan, 2010; Ouyang et al., 2018; J. Zhang, Kan, Shan, & Chen, 2016) and estimating length from morphometric relationships.

Machine vision could also be used to automate morphometric measurements which can be used to provide evidence of stock differentiation (Cadrin & Friedland, 1999; Rycroft, Radcliffe, & Atema, 2013). The manual extraction of measurements from images has been shown to be a reliable alternative to in-situ measurement (e.g. Rycroft et al., 2013) but it is time consuming. Automating the process could greatly increase throughput. The major technical difficulty will be to precisely and reproducibly identify landmarks. Combining whole body localization—as in the present study—with a detailed landmark search within the localized area using a feature detector (e.g. SIFT; Lowe, 2004) could be a way forward. Reducing the search space and to then reproducibly identify key points has been used successfully in face detection (Kazemi & Sullivan, 2014; Y. Sun et al., 2013) and open-source APIs exist (e.g. King, 2009).

6.5.5 Real-world deployment

Correction for lens distortion is critical for accurate photogrammetry as show in the present study, particularly with increased use of robust and waterproof action cameras (Claassens & Hodgson, 2018; Rogers et al., 2017; Schmid et al., 2017; Struthers et al., 2015) which have significant radial distortion. In small scale projects or where the camera model can be restricted then it may be practical for images to be undistorted on an ad hoc basis, for example by using the manufacturers own software. However, to deploy large scale volunteer based metrological data gathering it will be necessary to build a repository of lens correction profiles for each camera model where radial distortion is above an acceptable threshold. If a camera supports multiple focal lengths and field of views then these require separate profiles. Fortunately cameras typically embed state data (e.g. focal length) and camera model in image metadata which can be used to retrieve the correct profile to remove radial distortion. Profile creation is a relatively straight forward process from the photographer's / volunteer's point of view and involves taking multiple images of a regular pattern (e.g. a chessboard). For smartphones, profile creation could be embedded in the data gathering application itself and for non-smartphone cameras a web application could allow images to be submitted for profile creation. OpenCV (OpenCV team, 2018) provides the open-source code (used in the present study) to undistort images, although the API wrapper for the fish eye model was troublesome.

This article presents a closed problem with *a priori* knowledge that only a single class would occur in the image, this may not be unusual where interest is in a single species. CNNs are adept at discriminating between object classes (e.g. COCO, 2018, IMAGENET, 2018) and improved predictive models are frequently released as can be seen on Google's model zoo page (Google, 2018). The task of generalizing to additional species using R-CNN detectors and the combination of approaches outlined is eminently achievable for many species and CNNs have been used in fine grained species classification (e.g. Sun et al., 2016, Tamou et al., 2018).

In the present study good results were obtained with fewer than 1000 training images and this may be sufficient for fine grained species classification. CNNs have performed well in classifying images according to bird species with fewer than 100 examples per class (Lin, RoyChowdhury, & Maji, 2015). Nonetheless, data augmentation can be employed to improve the models (Ding, Chen, Liu, & Huang, 2016; Perez & Wang, 2017; Wong, Gatt, Stamatescu, & McDonnell, 2016). Augmentation transforms training images as part of the training pipeline to artificially boost the number of training images. Common transformations include rotation, blurring and elastic transformations, and CNN APIs usually have native support for augmentation. Alternatively augmentation can be managed prior to use in a preferred image processing API (e.g. Jung, 2018). It will be extremely difficult to use MV to discriminate between some species without large numbers of high resolution images. For example, identifying the flatfishes *Pleuronectes platessa*, *Limanda limanda* and *Platichtys flesus* is challenging even for postgraduate marine biologists (pers. observ.).

It will be impossible to obtain perfect object detections and length estimations, particularly in diary like volunteer applications. Pragmatically, users could be prompted to provide “hints” to any application to improve detection. For example, the IGFA fish catch log smartphone application (International Game Fish Association, 2018) prompts users to identify the snout and tail of the fish in an image to improve detection. This process could be used to determine subject rotation. Users could also be prompted to identify species where there may be uncertainty and these images can contribute to the training image set. Another smartphone application has used user contributed images to train a species classifier from submitted images (Fishbrain, 2018). Uncertain classifications and length estimations could be clarified by the general public by crowd sourcing as in other successful citizen science projects (e.g. Joly et al., 2014, Silvertown et al., 2015, Zooniverse, 2017) or by using paid-for crowdsourcing services (e.g. Amazon, 2017).

6.5.6 CONCLUSION

Collecting species and environmental data is a core task in marine and terrestrial ecology, and images are being used to monitor disease occurrence (e.g. Boesea et al., 2008, Barbedo, 2017), for species identification (Branson, Van Horn, Belongie, & Perona, 2014; Joly et al., 2014; Nilsback & Zisserman, 2008) and a host of other applications (Kühl & Burghardt, 2013). It is clear that images potentially encode much valuable data which is time consuming to process manually. CNNs are transforming image classification and object detection and excellent detection results can now be achieved from most images. Automatically extracting metrological data from images provides opportunities to greatly increase the volume and type of data that can be collected in many data gathering scenarios such as national citizen science programmes, directed surveys, remote electronic monitoring (e.g. camera traps), virtual observers with camera traps and other applications (review Bicknell et al., 2016). Further research is needed to reduce the potential bias and increase precision in extracted data in automated machine vision systems to achieve mainstream adoption, but continued advances in the technology will make machine vision approaches to data processing in ecology and fisheries an increasingly viable option without needing a computer science expert to develop MV solutions.

6.6 DATA ACCESSIBILITY

Code, Tensorflow configuration files, data and images with the ground truth rectangles defined in the VGG Image Annotator (http://www.robots.ox.ac.uk/~vgg/software/via) are published at https://github.com/seabass-detection/seabass-detection. Training and object detection made use of the Tensorflow object detection API, available at https://github.com/tensorflow/models/tree/master/research/object_detection.

7 General Discussion

7.1 SOCIAL MEDIA

It was shown how social media (chapter 3, C3-TDM) and other local knowledge sources (chapter 4, C4-LK) can be used to derive meaningful quantitative and qualitative estimators of key descriptors of marine recreational fisheries. Data collation and processing was largely a desktop exercise however, the compilation of local knowledge may require visits to MRF organisations or individuals (e.g. clubs). The cost effectiveness of this desktop exercise makes it repeatable. Repeatability is further enhanced by the automation of the methodological approachs presented.

Using unsolicited fisher knowledge allows researchers and government agencies to gain insights into MRF without requiring the cooperation of fishers. However, this requires subtle consideration of the ethics surrounding the use of passively contributed data, as discussed in Chapter 2. It was shown that TDM of these passively contributed data provided many more data points than those collected during general on-site surveys. Although not presented within C3-TDM, it was evident that discussion forums also provided substantially more records of fisher trips than that held collectively by recreational fishers, professionals involved in recreational fishing and fisher organisations. These sources were contacted but little data were forthcoming and contact management and arranging personal meetings was time intensive when compared to TDM.

Where certain groups of fishers are reluctant to share information with others—as is common in consumptive wildlife recreation (eg Maurstad, 2002; Olsen & Thuen, 2013; Svensson, 2016)—then fisher knowledge may represent the only viable means of collecting data without using direct observation by human observers, drones or aerial surveys. However, this short term gain has the potential to damage the fragile trust between academic researchers and government agencies. This reputational damage could influence people who share social networks (Bond et al., 2012; Centola, 2010) and be a barrier to the acceptance of science (Wynne, 1992). Sources can be anonymised, and this approach is recommended for reports, but may conflict with the standards required for scientific publication where methods should be transparent (accepting the anonymity of individual subjects).

Ostensibly, determination of the spatial and temporal proxies for MRF activity in C3-TDM and C4-LK may appear similar. Indeed georeferenced locations with recorded activity in

C3-TDM were one of the coverages used in C4-LK. Nevertheless, C4-LK used the frequency of sites mentioned in sources as a spatially referenced effort index from which temporal information or any effort metrics would be impossible to derive. The limitations of the C4-LK approach in providing detailed descriptors of activity are apparent, but where social media data is limited or non-existent it may represent the only passive reproducible methodology to obtain spatial effort indices. Expert opinion has been used in the past (e.g. Pawson et al., 1987) but this is not a repeatable methodology.

Only a limited number of the possible quantitative estimators which could have been extracted were presented in the C3-TDM paper. Extensive length and weight estimates of sea bass were also extracted however, these were not published as there was an expectation that this would cause an adverse reaction amongst online communities who can be hostile to use of their public data (Nardi, 2015). Such community hostility would have been unfortunate as a national angling survey was underway. Catches per trip were also recorded, hence these data could be used to analyse signals of declining size and changes in population structure (e.g. Bellquist and Semmens, 2016; Richardson et al., 2006) and decreasing numbers by species (e.g. McClenachan, 2009). The spatial distribution of effort in C3-TDM was aggregated to a high spatial area. Data were aggregated after considering the sensitivities of the MRF communities, some of which have an expectation that the locations they share within their online community are somehow private. It was shown that SM data can produce time series of quantitative effort and catch indices which are more detailed than the ordinal indices of the approach of C4-LK.

Where sufficient data are available, the C3-TDM method yields richer data which could be used to derive stock abundance indices, partial population structures (catchability and trophy reporting will create biases), and temporospatial distributions of CPUE. As extensively discussed in C3-TDM, these will be subject to unknowable and potentially very large biases however, it could be argued that such estimators are better than records of zero recreational catches (Pauly, 1998). Indeed, population estimators of total catch have been derived using social media or local knowledge data to provide the only description of MRF activity at a national level (Belhabib et al., 2016; N. S. Smith & Zeller, 2015). The richer data of C3-TDM—assuming sufficient data—would be better than the C4-LK method. The C3-TDM methodology can differentiate effort and catch by species, which is important in meeting MSFD obligations. It would be interesting to expand the C3-TDM methodology to include all species and validate the spatial distribution of effort against the FMM and WAM surveys introduced in C4-LK. It is also interesting that standardisation by shore length did not result in better prediction of the spatial distribution of effort in C4-LK, despite standardisation by shore length in C3-TDM. The focus of C3-TDM on temporal effort and the spatial aggregation of results may have masked the indifferent performance of standardisation by shore length seen in C4-LK. In terms of the MSFD, standardisation of activity by shore length may unnecessarily obscure the more easily interpretable absolute effort index at each location.

The TDM methodology is dependent on RFs publishing large volumes of data to public SM. Where volumes are low and if no methodological repetition is anticipated then it would be more efficient to manually extract data. The initial scraping was greatly facilitated by the hierarchical structure of forums however, evidence suggested that the numbers who are using older social media platforms are falling, as demonstrated by the closure of several of sources used in C3-TDM. This can also be inferred from the ever increasing dominance of the global SNS giants (e.g. Facebook and Twitter). It is also inferred that as older MRFs leave the online communities through life’s attritions and replaced by new generations, then the dominance of the global SNS giants will accelerate. SNSs do have structures which replicate the discussion forum format, on which much catch data is reported however, users are not anonymous and informed consent from each individual may be required to use the data for scientific purposes, particularly if there is any retention of user details. Aside from the difficulty of seeking informed consent, user behaviour may change if they know they are being observed (Babbie, 2013).

7.2 IMAGES

The accuracy (-0.1% mean bias error) of the fully corrected machine vision TL estimates presented in Chapter 7 (C6-MV) was only made possible by the preliminary work presented in Chapter 6 (C5-FID). The strength of the C5-FID method is that it is entirely mechanistic and can be adapted to morphological measurement of length of any comparatively large organism. It is also applicable to any fiducial marker.

Unfortunately, achieving the highly accurate TL estimates using machine vision in C6-MV was not entirely mechanistic and required modelling of error under rotation and outlier removal. Although volunteers could be told to ensure their images frame the fish horizontally, this feels unsatisfactory and several areas of further work were suggested in the chapter. Some combination of the listed approaches likely to provide improvements in the accuracy of length estimation are as follows.

(i) Retraining of R-CNNs by augmenting the training set with randomly rotated images (discussion C6-MV).
(ii) Brute force detection in "unknown" images by incremental rotation, the best fitting estimate will maximise the width to height ratio of the detection box (discussion C6-MV).
(iii) The correction model may be generalizable to all fusiform fish, this could be validated, and additional models developed for fish shapes where the model does not produce good bias corrections of TL.
(iv) Produce a satisfactory mechanistic model which quantitatively relates the detection rectangle to the total length according to the limits of the detection rectangle under rotation.

It is difficult to overstate the potential utility of machine vision in the collection and processing of data in commercial and recreational fisheries assessments. The recent development of R-CNNs has made the techniques accessible to marine researchers and other stakeholders. One of the primary drivers behind C6-MV is to communicate to other marine researchers how accessible these extremely complex techniques have become by using APIs developed by MV researchers and computer scientists using a real-world example, in an important and contentious high-profile species.

There are numerous examples throughout this thesis of articles where images (and video) encode information on the fish stocks which recreational and commercial fishers prosecute, and on the nature of their activity. Here the detection problem had limited scope but with transfer learning R-CNN models can be trained to detect other objects (e.g. other species, boats and gear). Catch records are visually embedded within images published to social media and media sharing networks, and present in sources of fisher knowledge. These images have been manually analysed to derive many descriptors (effort, population structure, harvest, illegal activity, invasive species) of recreational fisheries (Banha, Veríssimo, Ribeiro, & Anastácio, 2017; Belhabib et al., 2016; Giovos et al., 2018; McClenachan, 2009; Rizgalla et al., 2017; Shiffman et al., 2017). Fisheries scientists and managers have been working with images to monitor commercial fisheries for decades (D. C. Bartholomew et al., 2018; Chang et al., 2010, 2009; Hold et al., 2015; Needle et al., 2015; Pasco et al., 2009; Strachan, 1993; van Helmond et al., 2017; White et al., 2006) and the advances in MV and hardware means angler apps have been able to classify images (Fishbrain, 2018; International Game Fish Association, 2018). The FishBrain angler app team is working on length estimation from opportunistic fiducial markers

(KH, pers. comm.). This thesis—to the best knowledge of the author—represents the only length estimation of a fish using a foreground fiducial marker, in photographs captured in the real-world conditions which would be encountered aboard a boat or on the shore.

Perhaps the most pertinent application for MV length estimation in MRF assessments is in web and angler apps. Apps can collect data in several roles; (i) as commercial apps marketed to add value to the angling experience (provided collected data is shared by benevolent commercial companies); (ii) as part of the diary phases of directed surveys; and (iii) as part of longer term citizen science style data collection programs. This is an extremely promising way to improve recruitment and retention of RFs to data collection efforts and has worldwide relevance with smartphone ownership being comparatively high even in emerging countries (Pew Research Center, 2018; Poushter, 2016). Smartphone images also benefit from a wealth of metadata which smartphones record and which is relevant to assessments of marine recreational fisheries (ICES, 2017d; Venturelli et al., 2017). Another obvious application is automatically processing images captured during remote observations of charter boats and private recreational fishing boats who volunteer to record diaries.

All smartphones allow voice recording and both Apple and Google provide APIs which allow developers to interact with microphone input on their smartphone operating systems. Certainly combining voice and image recognition with standard interface interactions in apps would provide the greatest opportunity to accurately record data while further reducing application interface interactions with users. MRFs would particularly benefit from any reduction in the number of physical interactions they have with their device because of salt water ingress, dirty and cold hands or simply being too busy. Passive acoustic observations are being used to produce effort indices for marine recreational boat users by installing acoustic sensors at choke points (Hyder and Vieira, pers. comms.). Results are equivalent to those provided by visual remote electronic monitoring (e.g. Keller et al., 2016; Lancaster et al., 2017; Parnell et al., 2010). The machine learning techniques applied to image analysis can also be applied to the analysis of acoustic data and the convolution neural network architectures used in this image analysis can also be used in acoustic classification (Hershey et al., 2017). With sufficient training data it should be possible to identify specific boats—as opposed to boat classes.

In the collection of fisheries dependent data in commercial fisheries, REM can be cheaper than using onboard observers without automating image analysis (Chang et al., 2010; National Oceanic and Atmospheric Administration, 2015a), although this may not always be the case

(National Oceanic and Atmospheric Administration, 2015b). Cost estimate differences between the two NOAA assessments were attributed to reduced discard reporting requirements in the pelagic fishery. It is notable that the major costs in the NOAA pelagic assessment attributed 40% and 44% of the total annual cost estimate to data processing and program management respectively. In the observer scenario, the two largest costs were program management (62%) and observers (29%). This strongly suggests that as machine vision techniques are refined and become more accessible to fisheries researchers then further cost reductions can be realised. It is suggested that this is equally applicable to any application which collects images for monitoring purposes. Examples include aerial monitoring (e.g. Veiga et al., 2010; Vølstad et al., 2006), monitoring shore and boat based RFs from fixed cameras (Greenberg & Godin, 2015; Smallwood et al., 2012; van Poorten et al., 2015) and recording slipway and harbour use by RF vessels (Hartill et al., 2016). In all sampling settings, automating length estimates can reduce self-sampling biases and also reduce the perception of researchers that measurements reported by non-scientists are biased or not recorded as rigorously as they are by trained observers or researchers (Kraan et al., 2013).

The provision of CS and volunteer based programs by researchers and government agencies encourages a two-way dialogue with MRFs. This is important to build trust, which can be difficult to establish with MRF groups (pers. observ.; Olsen and Thuen, 2013; Svensson, 2016). Trust between scientists, managers and fishers is likely to help establish effective co-management of fisheries (Armitage et al., 2009) and support the continuation of any data collection programs. Engagement by MRFs in data collection is also likely to raise the profile of the activity which can be overlooked amongst competing activities (K. St. Martin & Hall-Arber, 2008). This could provide MRF participants with stronger representation when conflicts between marine sectors need to be resolved in planning and management processes. Improving the tools used by MRFs in volunteer programs should increase uptake and decrease volunteer dropout.

Using novel data sources to reconstruct recreational fishery and targeted stocks is problematic, but attempts have been made (Schiller et al., 2013, 2015; N. S. Smith & Zeller, 2015; D. Zeller et al., 2011; Dirk Zeller et al., 2007). A fundamental question is what degree of uncertainty and error in reconstructed time series will improve population estimators, or ultimately lead to better decision making than using no data (i.e. an estimate of zero RF induced mortality). Where some survey data has been intermittently collected, some interpolation of the missing points in the time series would appear to be a valid approach. In the absence of any data from an organised survey, then inferring trends observed in similar marine recreational fisheries could be used to project backwards from recently determined baselines. A purely hypothetical example for illustration, would be to take the current baseline of total annual effort for the UK and use relative changes in effort from the USA MRIP survey to make past predictions—noting that there are historical estimators of annual effort for the UK (e.g. NERC, 1970). Catch composition could be reconstructed from local ecological knowledge, interviews and other fisher knowledge records, and biomass estimated from those same records. Further adjustments could be made to the estimates based on our current understanding of the direction and magnitude of biases. It is accepted that this is an extremely simplistic picture. Nonetheless, as Pauly wrote: *The key part [to reconstruction] is psychological: one must overcome the notion that "no information is available" … fisheries are social activities [which] throw large shadows"* (Pauly, 1998). Certainly, access to the huge volumes of open text and images along with the means to automate information extraction from these sources will provide increasing opportunities for fisheries researchers to "fill in the gaps". Pauly could not have foreseen these emerging opportunities in his 1998 article. Much research remains to understand the nature of biases in these data, and the problem may yet prove to be insoluble.

8 Appendices

8.1 CHAPTER 3

Appendix A. Social Media Content Mining Overview

Social network sites such as Facebook and Twitter facilitate access to user generated content (UGC) by documented APIs which are an important part of the business model of these sites. Access to an API provide third parties with a defined object model (an object may be a user profile or a single tweet for example) and the filters to retrieve objects of interest (e.g. a user's post). APIs simplify and standardize data access and support content filtering prior to download. This server side filtering greatly reduces the amount of potentially irrelevant data which would otherwise be retrieved. However, social networking sites generally require users to register with real details and group content is private, making access practically and ethically more challenging.

Access to discussion forums and other social publishing mediums (e.g. blogs) is usually public, although a pseudonymous registration may be required. Discussion forums do not usually expose a public API, hence content is scraped via a standard page request to the web server, exactly equivalent to a person accessing the page with a web browser. The page structure of discussion forums and social publishing networks differs by individual sites however, a common approach to scraping content is used across different forums. Figure 8-1 outlines the content mining process, differentiating between social media sites which provide an API and those which do not.

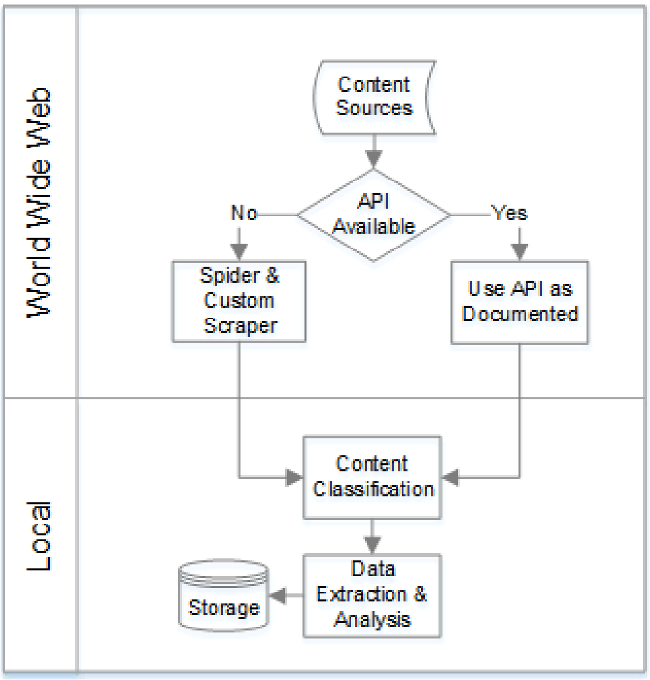

Figure 8-1. Idealised overview of social media content mining process. Examples of content sources include discussion forums, social network sites (e.g. Twitter) and visual and audio media focused sites (e.g. Flickr, YouTube). Social networking sites tend to provide application programming interfaces for developers to interact with content, whereas discussion forums require customized spiders (web crawlers) and scrapers. Local indicates that processing occurs on private computing resources, such as institutional servers (but can be cloud based services). Functional boundaries are becoming increasingly blurred, with provider APIs supporting domain/genre classification.

This outlined approach focuses on content mining from discussion forums but is generally applicable to other forms of social publishing. The process can be split into five stages as follows.

(1) Identification of social media sites amenable to content mining.

(2) Design a web crawler to navigate pages in the identified sites, or use an API (e.g. Facebook's Graph API).

(3) Download (scrape) text content and metadata (e.g. publication date) matching defined structural and textual constraints. For example, test for certain text content which appear between *<div class=”myclass”></div>* HTML tags

(4) Classify and label content to determine relevancy to the domain of interest.

(5) Information extraction from classified text to produce structured data.

CONTENT SOURCE IDENTIFICATION

Google is the most popular search engine in the world (netmarketshare.com, 2017), undoubtedly academic content miners frequently use Google's web search service to identify data sources and reference materials. However, Google do not publish information on how search results are personalised, nor provide an opt-out from results personalisation. The consequence is that searches return inconsistent results, even where researchers publish search terms. Hannak et al. (2013) showed that personalisation is strongly attributed to the searcher's IP address (a proxy for geographic location) and whether the user is logged into Google when searching. Search engines are available which do not personalise results (e.g. DuckDuckGo.com) however, results will still change over time because of search index updates and the distributed infrastructure of the search servers. Such inconsistency could be considered problematic in research because ethical considerations mean content sources are not named which conflicts with the principle of open research data and reproducibility. Unfortunately there is no easy answer; Google and other search engines will remain an important means of identifying online content.

CRAWLING, FILTERING AND SCRAPING SOURCES

Web crawling is the process of navigating URL links within one or more websites to discover navigable links (a sitemap) and the content to which the links point. The crawler will usually determine the relevancy of content according to rules defined at design time (prefiltering). For example, does the page title (i.e. text appearing between the <title></title> tags) contain word aliases of our species of interest, or does the page's content meet a certain keyword frequency. The workflow is not rigid and will depend on many factors including the complexity of the task, degree of automation required (attended or unattended execution), if the task is reoccurring, volumes of data to scrape, the accuracy required and the software tools used to perform the crawl.

The most common approach (termed a focused crawl) uses a predefined URL list (compiled from the site identification step) which is submitted to the crawler. Some simple logic (or additional crawling) then builds fully qualified URLs pointing to subpages and content during crawler execution. A manual review of site URLs is usually sufficient to determine the logic necessary to program a crawler to visit the relevant site pages. Discussion forums are particularly amenable to this approach because of their rigid hierarchical structure, making a broad crawl unnecessary.

Figure 8-2 shows a typical discussion forum structure with example URL hierarchy. We see that the forum has a board called *board1*. If *board1* has 30 pages with each page containing 20 threads, then there are 600 threads to scrape, each having a unique URL. *Board1* (as an example of a typical forum) uses incrementing suffixes to generate URLs for new pages (e.g. /pg1, /pg2) and for each new thread published by a user (e.g. /pg1/1.html, /pg1/2.html). Therefore if we know the total number of threads of interest, it is trivial to build a list of URLs to each thread within each page. We then have a complete list which 'tells' the crawler where it can retrieve the content of each post. At this stage we can introduce any necessary prefiltering. Prefiltering is useful when there is a large amount of potentially irrelevant content and we wish to minimise impacts on websites or reduce the complexity of our content classification and information tasks.

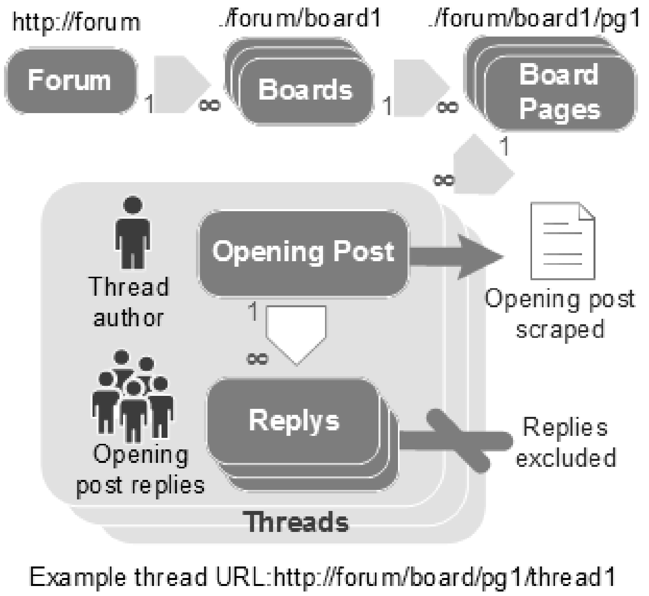

Figure 8-2. Hierarchical structure of a typical world wide web discussion forum. Symbols 1 and ∞ show a one-to-many relationship. Hence a forum has many subject boards, which has many pages containing many threads. A thread is created with an opening post. Only the content of the opening post was mined for this work.

Broad crawls can be performed where links and relevant content are automatically discovered across multiple domains and/or subdomains. Broad crawls would necessitate the introduction of complex classifiers to be fully automated or a manual review of content to compile a predefined URL list to perform a narrow crawl.

The penultimate stage of the scrape is the first pass parsing of UGC from the scraped content which can include script code, document mark-up (e.g. <title>), metadata and irrelevant content (e.g. advertising links). Parsing means extracting content between, following or after predefined delimiters which are usually document mark-up tags. Most tags will appear multiple times on a web page, but tag attributes can be used to identify specific instances (for example the *<TR>* tag is used to define a table row, but an id tag may uniquely identify the first table row: *<TR id=1>*). The definitive mark-up references are provided by the World Wide Web Consortium (W3C) which publishes HTML, CSS and XML mark-up language standards (W3C, 2017). Crawlers invariably support the XPath language, which can be used to precisely and easily (once XPath is mastered) extract data from any element in a web page's hierarchy. Following parsing, it may be necessary to clean content by removing unprintable characters or extraneous mark-up. The final scraping stage is to save the content to a data store under the researcher's control.

CLASSIFICATION AND INFORMATION EXTRACTION

For our purposes, classification is the assignment of meaning to parts of scraped content. Parts could be sentences within a thread or the thread as a whole. Meaning will be dependent on the research questions, for example, does a unit of text have a species observation, or record of wildlife-human interaction.

Computational approaches to classification are legion and potentially complex, we could not hope to scrape the topic surface. However, it is typical to split content into sentences and then perform classification on these sentences using machine learning classifier(s) implemented in a language of the researcher's choice. Sentences would be labelled according to the resulting classification(s), along with some metric indicating the strength of machine belief in the classification. As an example, we may wish to know if UGC indicated negative sentiment towards a policy banning fishing in spring. The classifier must assign a likelihood according to multiple criteria — that the sentiment applies to the policy and is negative, and that it is spring enforcement to which the content refers.

In its simplest form, classification is the assignment of labels simply indicating keyword presence according to a precompiled dictionary covering the lexicon used in the relevant domain of interest (the approach in the associated article). This simple approach will require considerable manual intervention in the extraction of information as sentence semantics are ignored. Where a simple approach is unsuitable, for example if there is a lot of UGC which is information sparse (e.g. in Twitter posts), or where automated real-time monitoring is the goal then probabilistic classifiers (e.g. naive Bayes) and deep machine learning algorithms (Murphy, 2012) can be trained. Supervised classification is the common approach, with an initial training data set (laboriously) classified by human experts. A machine learning algorithm then analyses the training set to create a predictive model for unattended machine classification. This predictive model is then used against scraped text to classify it.

The final stage is information extraction, classification allows the researcher to filter classified content by confidence thresholds on how likely content is to contain relevant meaning. Filtering can greatly increase the efficiency of data extraction by reducing the volume of text to be processed and 'cherry picking' well-formed content. Automated data extraction is a huge field and extensively covered in the natural language processing literature (e.g. Jurafsky & Martin, 2008). Briefly, it involves reducing sentences to their constituent words (tokenization) and then looking up those words in a corpora to assign root meanings (known as parts of speech tagging). Additional stages can attribute higher level sentence meanings according to the sentence structure and the context in which the sentence appears. This involves considerable time and effort to achieve robust results and will generally be unjustified for small projects.

Mixing manual (human) and machine learning for classification and extraction is a more realistic approach. However, citizen science and crowdsourcing can be viable approaches, particularly where receptive stakeholder communities exist. Wider crowd sourcing communities can also be engaged via online platforms such as Zooniverse (Zooniverse, 2017) which has many classification based projects. Even when stakeholders are unavailable, researchers can crowdsource via incentivization platforms such as Amazon's Mechanical Turk (Amazon, 2017b).

NOTES ON SOFTWARE

Separate software and APIs are used for scraping and classification/information extraction. Scraping software can be purchased 'off-the-shelf' and usually provides a graphical user interface (GUI) to allow content to be scraped and results to be viewed with no knowledge of programming and a rudimentary knowledge of web page structure. Such software usually supports unattended and scheduled scraping and mechanisms to reduce the load on targeted sites. For most scenarios these solutions will be more than adequate.

Where off-the-shelf scrapers are inadequate, then programming languages will almost certainly have open source libraries available. Researchers familiar with Python can use Scrapy (Scrapy, 2017) and those who prefer R have rvest (Wickham, 2016). These libraries are extensible and so the programmer can customise their solution according to any requirements not available in the core package. Paid-for Software as a Service (SaaS) providers are also available and provide scraping services which are executed from the provider's servers. SaaS providers also offer anonymised scraping and (ethically questionable) methods of preventing target websites from detecting and blocking scraping activity. Modern sites tend to use asynchronous techniques to load areas of a page dynamically, which makes scraping less intuitive. Specialist packages are available to scrape dynamic content (e.g. Selenium and asyncio).

There are too many machine learning APIs and SaaS like providers available to mention. Many packages are available in both R, Python, and specialised open source applications. The most mature NLP orientated packages are the OpenNLP (Hornik, 2016).and RWeka (Hornik et al., 2017) libraries for R and NTLK for Python (Bird, Loper, & Klein, 2009). Each supports core NLP functions (tokenization, parts of speech labelling etc.) and, at least, a maximum entropy classifier. SaaS based platforms are also available as paid-for services (e.g. Amazon, 2017a; Google, 2017; Microsoft, 2017) and are able to provide high performance computing power in an environment with both software and hardware support. However, the algorithms used by some SaaS providers are opaque, closed source and subject to change without notification.

Appendix B. Parse Functions

```
Imports System.Text

Module ParseFunctions


    Public Function CheckDistance(ByVal sentence As String, ByVal
first As String, ByVal second As String, ByVal distance As Integer,
Optional ByVal anyorder As Boolean = True) As Boolean
        Dim a$, b$

        a = "\b" + first + "\W+(?:\w+\W+){0," + distance.ToString +
"}?" + second + "\b"
        b = "\b" + second + "\W+(?:\w+\W+){0," + distance.ToString +
"}?" + first + "\b"

        Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
        Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)

        Dim m As RegularExpressions.Match = reFL.Match(sentence)
        Dim n As RegularExpressions.Match = reLF.Match(sentence)

        If anyorder Then
            CheckDistance = m.Success Or n.Success
        Else
            CheckDistance = m.Success
        End If
    End Function


    Public Function CheckDistanceTimeMultiText(ByVal sentence As
String, ByVal thetext As String(), ByVal distance As Integer,
Optional ByVal anyorder As Boolean = True) As Boolean
        Dim a$, b$
        For Each s As String In thetext
            a = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}?" +
s.ToString + "\b"
```

```
            b = "\b" + s.ToString + "\W+(?:\w+\W+){0," +
distance.ToString + "}(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\b"


            Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)


            Dim m As RegularExpressions.Match = reFL.Match(sentence)
            Dim n As RegularExpressions.Match = reLF.Match(sentence)


            If anyorder Then
                CheckDistanceTimeMultiText = m.Success Or n.Success
            Else
                CheckDistanceTimeMultiText = m.Success
            End If
            If CheckDistanceTimeMultiText = True Then Exit Function
        Next s
        CheckDistanceTimeMultiText = False
    End Function



    Public Function CheckDistanceAnyNumber(ByVal sentence As String,
ByVal thetext As String, ByVal distance As Integer, Optional ByVal
anyorder As Boolean = True) As Boolean
        Dim a$, b$


        a = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}?" +
thetext + "\b"
        b = "\b" + thetext + "\W+(?:\w+\W+){0," + distance.ToString
+ "}(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-9])?|0?\.\d*[1-
9]|0)\b"


        Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
        Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)


        Dim m As RegularExpressions.Match = reFL.Match(sentence)
```

```
        Dim n As RegularExpressions.Match = reLF.Match(sentence)

        If anyorder Then
            CheckDistanceAnyNumber = m.Success Or n.Success
        Else
            CheckDistanceAnyNumber = m.Success
        End If
    End Function

    Public Function CheckDistanceAnyNumberMultiText(ByVal sentence
As String, ByVal thetext As String(), ByVal distance As Integer,
Optional ByVal anyorder As Boolean = True) As Boolean
        Dim a$, b$

        For Each s As String In thetext
            a = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}?" +
s.ToString + "\b"
            b = "\b" + s.ToString + "\W+(?:\w+\W+){0," +
distance.ToString + "}(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\b"

            Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)

            Dim m As RegularExpressions.Match = reFL.Match(sentence)
            Dim n As RegularExpressions.Match = reLF.Match(sentence)

            If anyorder Then
                CheckDistanceAnyNumberMultiText = m.Success Or
n.Success
            Else
                CheckDistanceAnyNumberMultiText = m.Success
            End If
            If CheckDistanceAnyNumberMultiText = True Then Exit
Function
        Next s
        CheckDistanceAnyNumberMultiText = False
    End Function
```

```
    Public Function GetMatchedStringswithTime(ByVal sentence As String, ByVal thetext As String, ByVal distance As Integer, Optional ByVal anyorder As Boolean = True) As ArrayList
        Dim a$, b$
        Dim aa As New ArrayList
        '(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-9]$)|(^([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-5][0-9]$)

        a = "\b(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-9]$)|(^([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-5][0-9]$)\W+(?:\w+\W+){0," + distance.ToString + "}" + thetext + "\b"
        b = "\b" + thetext + "\W+(?:\w+\W+){0," + distance.ToString + "}(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-9]$)|(^([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-5][0-9]$)\b"

        Dim reFL As RegularExpressions.Regex = New RegularExpressions.Regex(a, RegularExpressions.RegexOptions.IgnoreCase)
        Dim reLF As RegularExpressions.Regex = New RegularExpressions.Regex(b, RegularExpressions.RegexOptions.IgnoreCase)

        Dim m As RegularExpressions.MatchCollection = reFL.Matches(sentence)
        Dim n As RegularExpressions.MatchCollection = reLF.Matches(sentence)


        'order is time-text only when anyorder is false
        If m.Count > 0 Then
            For Each match As RegularExpressions.Match In m
                aa.Add(match.Value)
            Next
        End If
        If anyorder Then
            If n.Count > 0 Then
                For Each match As RegularExpressions.Match In n
                    aa.Add(match.Value)
                Next
            End If
        End If
        GetMatchedStringswithTime = aa
    End Function
```

```
    Public Function GetMatchedStringswithTimeMultiText(ByVal
sentence As String, ByVal thetext As String(), ByVal distance As
Integer, Optional ByVal anyorder As Boolean = True) As ArrayList
        Dim a$, b$
        Dim aa As New ArrayList
        '(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-9]$)|(^([0-9]|0[0-
9]|1[0-9]|2[0-3]).[0-5][0-9]$)

        For Each s As String In thetext
            a = "\b(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-
9]$)|(^([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-5][0-9]$)\W+(?:\w+\W+){0," +
distance.ToString + "}" + s.ToString + "\b"
            b = "\b" + s.ToString + "\W+(?:\w+\W+){0," +
distance.ToString + "}(^([0-9]|0[0-9]|1[0-9]|2[0-3]):[0-5][0-
9]$)|(^([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-5][0-9]$)\b"

            Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)

            Dim m As RegularExpressions.MatchCollection =
reFL.Matches(sentence)
            Dim n As RegularExpressions.MatchCollection =
reLF.Matches(sentence)



            'order is time-text only when anyorder is false
            If m.Count > 0 Then
                For Each match As RegularExpressions.Match In m
                    aa.Add(match.Value)
                Next
            End If
            If anyorder Then
                If n.Count > 0 Then
                    For Each match As RegularExpressions.Match In n
                        aa.Add(match.Value)
                    Next
                End If
            End If
        Next s
```

```
        GetMatchedStringswithTimeMultiText = aa
    End Function


    Public Function HasMatchedStringswithTimeMultiText(ByVal
sentence As String, ByVal thetext As String()) As Boolean
        Dim a$
        For Each s As String In thetext
            a = "(?=.*\b" + s + "\b)((?=.*\b(([0-9]|0[0-9]|1[0-
9]|2[0-3]):[0-5][0-9])\b)|(?=.*\b(([0-9]|0[0-9]|1[0-9]|2[0-3]).[0-
5][0-9])\b))"

            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)
            If m.Success Then
                HasMatchedStringswithTimeMultiText = True
                Exit Function
            End If
        Next s
        HasMatchedStringswithTimeMultiText = False
    End Function



    Public Function GetMatchedStringsAnyNumber(ByVal sentence As
String, ByVal thetext As String, ByVal distance As Integer, Optional
ByVal anyorder As Boolean = True) As ArrayList
        Dim a$, b$
        Dim aa As New ArrayList
        '(?:[1-9](?:\d{0,2})(?:,\d{3})*(?:\.\d*[1-9])?|0?\.\d*[1-
9]|0)
        '\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0,10}bass\b

        a = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}" +
thetext + "\b"
        b = "\b" + thetext + "\W+(?:\w+\W+){0," + distance.ToString
+ "}(?:[1-9](?:\d{0,2})(?:,\d{3})*(?:\.\d*[1-9])?|0?\.\d*[1-9]|0)\b"

        Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
```

```
        Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)

        Dim m As RegularExpressions.MatchCollection =
reFL.Matches(sentence)
        Dim n As RegularExpressions.MatchCollection =
reLF.Matches(sentence)


        'order is number-text only when anyorder is false
        If m.Count > 0 Then
            For Each match As RegularExpressions.Match In m
                aa.Add(match.Value)
            Next
        End If
        If anyorder Then
            If n.Count > 0 Then
                For Each match As RegularExpressions.Match In n
                    aa.Add(match.Value)
                Next
            End If
        End If
        GetMatchedStringsAnyNumber = aa
    End Function


    Public Function GetMatchedStringsAnyNumberMultiText(ByVal
sentence As String, ByVal thetext As String, ByVal otherWords As
String(), ByVal distance As Integer) As ArrayList
        Dim a$, b$, c$, d$
        Dim aa As New ArrayList
        '(?:[1-9](?:\d{0,2})(?:,\d{3})*(?:\.\d*[1-9])?|0?\.\d*[1-
9]|0)
        '\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0,10}bass\b
        sentence = " " + sentence.ToLower + " " 'looks odd, but
        thetext = thetext.ToLower
        For Each s As String In otherWords
            s.ToLower()
            'these 4 permutations are sufficient for english
sentence constructs "5 bass of a pound","5 bass of 5 pounds", 10
```

```
pound bass, "bass in pounds, was 10", bass of 10 pounds, "a 5 pound
bass"
            'n bass pound
            a = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}(" +
thetext + ")" + "\W+(?:\w+\W+){0," + distance.ToString + "}(" + s +
")\b"


            'bass pound n
            b = "\b(" + thetext + ")\W+(?:\w+\W+){0," +
distance.ToString + "}(" + s + ")\W+(?:\w+\W+){0," +
distance.ToString + "}(?:[1-9](?:\d{0,2})(?:,\d{3})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\b"


            'bass n pound
            c = "\b(" + thetext + ")\W+(?:\w+\W+){0," +
distance.ToString + "}(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}(" + s
+ ")\b"


            'n pound bass
            d = "\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\W+(?:\w+\W+){0," + distance.ToString + "}(" + s
+ ")" + "\W+(?:\w+\W+){0," + distance.ToString + "}(" + thetext +
")\b"


            Dim reFL As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reLF As RegularExpressions.Regex = New
RegularExpressions.Regex(b,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reMID As RegularExpressions.Regex = New
RegularExpressions.Regex(c,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim reNPB As RegularExpressions.Regex = New
RegularExpressions.Regex(d,
RegularExpressions.RegexOptions.IgnoreCase)


            Dim m As RegularExpressions.MatchCollection =
reFL.Matches(sentence)
            Dim n As RegularExpressions.MatchCollection =
reLF.Matches(sentence)
            Dim o As RegularExpressions.MatchCollection =
reMID.Matches(sentence)
            Dim p As RegularExpressions.MatchCollection =
reNPB.Matches(sentence)
```

```
            If m.Count > 0 Then
                For Each match As RegularExpressions.Match In m
                    If Not aa.Contains(match.Value) Then
aa.Add(match.Value)
                Next
            End If

            If n.Count > 0 Then
                For Each match As RegularExpressions.Match In n
                    If Not aa.Contains(match.Value) Then
aa.Add(match.Value)
                Next
            End If

            If o.Count > 0 Then
                For Each match As RegularExpressions.Match In o
                    If Not aa.Contains(match.Value) Then
aa.Add(match.Value)
                Next
            End If

            If p.Count > 0 Then
                For Each match As RegularExpressions.Match In p
                    aa.Add(match.Value)
                Next
            End If
        Next s
        GetMatchedStringsAnyNumberMultiText = aa
    End Function


    Public Function SentenceIsMatchAndHasNumeric(ByVal sentence As
String, ByVal thetext As String, ByVal otherWords As String()) As
Boolean
        Dim a$
        For Each s As String In otherWords
            a = "(?=.*\b" & thetext & "\b)(?=.*\b" & s.ToString &
"\b)(?=.*\b(?:[1-9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-
9])?|0?\.\d*[1-9]|0)\b)"
```

```
            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)

            If m.Success Then
                SentenceIsMatchAndHasNumeric = True
                Exit Function
            End If
        Next s
        SentenceIsMatchAndHasNumeric = False
    End Function

    Public Function SentenceHasTextAndNumber(ByVal sentence As
String, ByVal thetext As String) As Boolean
        Dim a$
        a = "(?=.*\b" & thetext & "\b)(?=.*\b(?:[1-
9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-9])?|0?\.\d*[1-9]|0)\b)"

        Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
        Dim m As RegularExpressions.Match = rex.Match(sentence)

        If m.Success Then
            SentenceHasTextAndNumber = True
            Exit Function
        End If

        SentenceHasTextAndNumber = False
    End Function

    Public Function SentenceHasTextMultiAndNumber(ByVal sentence As
String, ByVal thetext() As String) As Boolean
        Dim a$
        For Each s As String In thetext
            a = "(?=.*\b" & s & "\b)(?=.*\b(?:[1-
9](?:\d{0,10})(?:,\d{10})*(?:\.\d*[1-9])?|0?\.\d*[1-9]|0)\b)"

            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)
```

```
            If m.Success Then
                SentenceHasTextMultiAndNumber = True
                Exit Function
            End If
        Next
        SentenceHasTextMultiAndNumber = False
    End Function

    Public Function SentenceHasTextMulti(ByVal sentence As String,
ByVal thetext() As String) As Boolean
        Dim a$
        For Each s As String In thetext
            a = "(?=.*\b" & s & "\b)"
            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(a,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)

            If m.Success Then
                SentenceHasTextMulti = True
                Exit Function
            End If
        Next s
        SentenceHasTextMulti = False
    End Function

    Public Function GetSurroundingWords(ByVal sentence As String,
ByVal theText As String, ByVal nr As Integer) As String
        Dim s As String = " "
        Dim parts As String() = sentence.ToLower.Split(CChar(" "))

        For i As Integer = 0 To parts.Length - 1
            If parts(i) = theText.ToLower.Trim Then
                s = theText & " "
                For x As Integer = -1 * nr To i - 1
                    If x + i >= 0 And x + i <= UBound(parts) Then s
= s + parts(x + i) + " "
                Next
                For x = i + 1 To i + nr
                    If x <= UBound(parts) Then s = s + parts(x) + "
"
```

```
                Next
            End If
        Next
        s = s.Trim
        GetSurroundingWords = s
    End Function

    Public Function GetSurroundingWordsMultiMatches(ByVal sentence
As String, ByVal theText As String(), ByVal nr As Integer) As String
        Dim s As String = " "
        Dim parts As String() = sentence.ToLower.Split(CChar(" "))

        For i As Integer = 0 To parts.Length - 1
            For z As Integer = 0 To UBound(theText)
                If parts(i) = theText(z).ToLower.Trim Then
                    s = theText(z) & " "
                    For x As Integer = -1 * nr To i - 1
                        If x + i >= 0 And x + i <= UBound(parts)
Then s = s + parts(x + i) + " "
                    Next
                    For x = i + 1 To i + nr
                        If x <= UBound(parts) Then s = s + parts(x)
+ " "
                    Next
                    s = s.Trim
                    GetSurroundingWordsMultiMatches = s
                    Exit Function
                End If
            Next
        Next
        GetSurroundingWordsMultiMatches = sentence
    End Function

    Public Function TokenInString(ByVal sentence As String, ByVal
strArray As String()) As Boolean
        Dim nlp As New nlp
        Dim tokens As String() =
nlp.TokenizeSentence(sentence.ToLower)
        For Each token As String In tokens
            token = token.Trim
            For Each mystr As String In strArray
                If token = mystr.Trim Then
```

```
                    TokenInString = True
                    Exit Function
                End If
            Next mystr
        Next token
        TokenInString = False
    End Function

    Public Function StringArrayInSentenceGetFirst(ByVal sentence As
String, ByVal al As String(), Optional ByVal check_whole_word As
Boolean = True) As String
        Dim tmp$
        sentence = sentence.ToLower
        For Each s As String In al
            If check_whole_word Then tmp = "\W" & s.ToLower.Trim &
"\W" Else tmp = s.ToLower.Trim
            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(tmp,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)
            If m.Success Then
                StringArrayInSentenceGetFirst = s
                Exit Function
            End If
        Next
        StringArrayInSentenceGetFirst = ""
    End Function



    Public Function StringArrayInSentenceGetAll(ByVal sentence As
String, ByVal al As ArrayList, Optional ByVal check_whole_word As
Boolean = True) As ArrayList
        Dim tmp$
        Dim alRet As New ArrayList
        sentence = sentence.ToLower
        For Each s As String In al
            If check_whole_word Then tmp = "\W" & s.ToLower.Trim &
"\W" Else tmp = s.ToLower.Trim
            Dim rex As RegularExpressions.Regex = New
RegularExpressions.Regex(tmp,
RegularExpressions.RegexOptions.IgnoreCase)
            Dim m As RegularExpressions.Match = rex.Match(sentence)
            If m.Success Then alRet.Add(s)
```

```
        Next
        StringArrayInSentenceGetAll = alRet
    End Function
End Module
```

Appendix C. Activity and Intensity Calculation Details

MONTHLY ACTIVITY PATTERNS

The number of posts to social media increased from March 2002 because of increasing user adoption, hence it is necessary to standardise effort estimates derived from social media posts. Time series stationarity is also important for the statistical tests used. This section describes how time series were created and standardized from mined data and details how the following reported metrics were calculated.

(iv) $\{M\}_{|140|}$[**], time series of unadjusted *monthly activity* (gear hours) between March 2002 and September 2013.

(v) $\{Z\}_{|140|}$, time series of *standard monthly activity*. The monthly activity $\{M\}$ differenced by the mean monthly activity for a year, divided by the standard deviation of mean monthly activity for the year.

(vi) $\{\acute{Z}\}_{|140|}$, time series of *y-j standard monthly activity*. The Yeo-Johnson transform of $\{Z\}$, homoscedastic.

Expressing formally, let $\boldsymbol{m}_{\{1,\, 2, \dots\, 140\}}$ be the months between March 2002 and September 2013, let $\boldsymbol{y}_{\{2002,\, 2003,\, \dots\, 2013\}}$ be the years from 2002 to 2013. Let $\boldsymbol{x}$ denote the sampling variable of gear hours per mined trip. Then, $\{M\}_m = \sum_{x_i \in m} x_i$ [††] where. To derive $\{Z\}$ and $\{\acute{Z}\}$, let $\{Z\}_{|140|}$ be the time series calculated according to (Eq. 1). Finally, $\{\acute{Z}\} = \varphi(\lambda = 0.4, \{Z\})$, which represents the Yeo-Johnson power transform (Fox & Weisberg, 2011; Yeo & Johnson, 2000).

(1) $$\{\bar{A}\}_y = \frac{1}{n} \cdot \sum_{m \in y}^{n} \{M\}_m, \qquad \text{then } \{Z\}_m = \frac{\{M\}_m - \{\bar{A}\}_y}{SD\big(\{\bar{A}\}_y\big)} \;\mid y_k \ni m^{\ddagger\ddagger}$$

SEA ANGLING 2012 MEAN MONTHLY ACTIVITY

Trip durations were calculated by summing the hours fished and hours still to fish reported by respondents during face-to-face interviews (Armstrong et al., 2013a; CEFAS, 2012). To compare the Sea Angling 2012 (SA2012) survey data with $\{\acute{Z}\}$, all variables required mean aggregation to month. The following variables are reported in the results.

(iv) $\bar{\acute{Z}}_{|12|,\, mean}$ *y-j standard monthly* activity from the time series $\{\acute{Z}\}$.

[**] The vertical bar notation (e.g. |140|) is the number of elements (cardinality) of an array, **not** the modulus function.
[††] Here, the general notation $x_i \in y$ denotes that x_i belongs to y_i, e.g. for $x_i \in m_j$, x_i occurred in month m.
[‡‡] Here, $y_j \ni m$ denotes that the indexed year y is the year which temporally contains the month m.

(v) $\bar{Z}_{sa_{cod}}$|12| standardized (Z) scores of total monthly trip hours from SA2012 survey trips with a recorded cod capture.

(vi) $\bar{Z}_{sa_{bass}}$|12|, as $\bar{Z}_{sa_{cod}}$, for trips with a recorded bass capture.

Formally, let $k_{\{\text{Jan, Feb, ..., Dec}\}}$ be the set of months. Then $\bar{\acute{Z}}$is calculated according to Eq. 2. Let $\bar{Z}_{sa}$ represent either $\bar{Z}_{sa_{cod}}$ or $\bar{Z}_{sa_{bass}}$, Let $\boldsymbol{x}$ denote SA2012's sampled variable trip duration, let s_{sa} be SD(s_k) where $s_k = \sum_{x_i \in k}^{k} x_i$, and let the monthly effort adjustment factor $a_k = {}^{12 \cdot \sum_{d_i \in k} d_i}\!/\!_{\sum d_i}$ where d_i is the variable of survey visit durations, then $\bar{Z}_{sa}$, is given by Eq. 3.

(2) $$\bar{\acute{Z}}_k = \frac{1}{n} \cdot \sum_{\{\acute{Z}\}_m \in k}^{n} \{\acute{Z}\}_m$$ [§§]

(3) $$\bar{Z}_{sa_k} = \frac{\frac{1}{a_k} \cdot \sum_{x_i \in k} x_i - \frac{1}{12} \cdot \sum x_i}{s_{sa}}$$

SEASONAL AND SPATIAL ACTIVITY PATTERNS

This section outlines how the following reported metrics were calculated.

(iv) $\bar{\boldsymbol{x}}_{\boldsymbol{season}}$ |2| Mean seasonal activity (gear hours region^{-1} year^{-1}), unadjusted mean of gear hours for season.

(v) $\mathbf{I}_{sr}$ |24|, Intensity (gear hours km^{-1} month^{-1}), standardised by coastline length, stratified by summer and winter across the 24 regions to give a total of 48 strata.

(vi) $\mathbf{D}_{sr}$ |24|, Differenced intensity (gear hours km^{-1} month^{-1}), $\mathbf{I}_{sr}$ (see [ii]) differenced using yearly intensity means.

Spatial regions were created using a 25 km by 25 km grid, intersected with the 6 nautical mile national limit and the Wales mean high water mark (1:50 0000 scale). Polygons under ~50 km^2 were merged with the neighbour sharing the most similar marine area characteristics according to Parker (2015) to give 24 regions. Shore length was considered more indicative of 'available opportunities to fish' than area, and high water shore length more appropriate than low water length. To control for variations in the topological complexity of the coastline, a polynomial approximation with exponential kernel (PAEK) smoothing with 500 m tolerance

[§§] The term $\{\acute{Z}\}_m \in k$denotes that the indexed months *m* is a member of calendar month *k* was applied. The output was reviewed for locations known to the authors, to validate the removal of ‘meso level’ shore features while preserving > 50 m features. The resulting smoothed coastal length for Wales was 2032 km.

All calculations were derived from the trended time series $\{M\}$ above. $\{M\}$ was trend stationary from 2006 to 2013 (Mann-Kendall Hamed Rao, p = 0.07) and sample numbers between 2002 and 2005 (n = 2, 2, 8, 23 respectively) were low, hence $\{M\}$ was truncated to elements from 2006 to 2013 ($\{\breve{M}\}$). Visually, $\{\breve{M}\}$ appeared non-stationary (e.g. a perceptible decreasing trend after 2011), hence a within-year standardisation was applied, as detailed below. Let $\boldsymbol{r}_{\{1,2,\ldots 24\}}$ be the 24 regions , let $\boldsymbol{y}_{\{2006, 2007, \ldots 2013\}}$ be the years, and let $\boldsymbol{s}_{\{\text{summer, winter}\}}$ be the seasons to which elements of $\{\breve{M}\}$ belong. The mean seasonal activity, $\bar{x}_{season}$is given in Eq. 4.

$$\bar{x}_{season} = \frac{1}{|y| \cdot |r|} \cdot \sum_{\{\breve{M}\}_i \in s} \{\breve{M}\}_i \tag{4}$$

Let l_r = smoothed shore length of region r, then the intensity matrix $\mathbf{I}_{sr}$ was calculated according to Eq. 5 and the differenced intensity matrix $\mathbf{D}_{sr}$ was calculated according to equation Eq. 6.

$$I_{sry} = \frac{1}{n \cdot l_r} \cdot \sum^{n}_{\substack{\{\breve{M}\}_i \in s \,\cap \\ \{\breve{M}\}_i \in r \,\cap \\ \{\breve{M}\}_i \in y}} \{\breve{M}\}_i, \qquad \text{and let } \mathbf{I}_{sr} = \bar{I}_{sry} \tag{5}$$

$$D_{sry} = I_{sry} - \frac{1}{\mathrm{n} \cdot l_r} \cdot \sum^{\mathrm{n}}_{\substack{\{\breve{M}\}_i \in r \,\cap \\ \{\breve{M}\}_i \in y}} \{\breve{M}\}_i, \qquad \text{and let } \mathbf{D}_{sr} = \bar{D}_{sry} \tag{6}$$

Appendix D. Compiling the Lexicon

A lexicon of terms was compiled so that opening forum posts and the sentences of opening posts can be labelled with the type of data they contain. Broadly, the aim is to maximise the number of correct classifications per unit volume of text while minimising false positives. In addition, if the aim is to process large volumes of data, (e.g. in real time monitoring) then randomised sampling from relevant social media should be considered to avoid potential bias. An example would be the use of different colloquial names for the species studied by geographical location. In the present study, several approaches were taken to compilation of terms. A brief description some of those methods and others which could be considered follows.

EXPERT AND PARTICIPANT VOCABULARY

Participants engaging in wildlife recreation will be familiar with the colloquial terms used by participants. Hence participant and expert knowledge (e.g. wildlife managers) can identify words and phrases not found in formal lexical sources. The authors' conducted unstructured face to face interviews with three marine recreational fishers who prosecute bass from the shore. These fishers volunteered common recreational gear use and colloquial names for seabass. It is important to talk to participants as colloquial terms for juvenile and mature specimens, or for male and females could be common.

LEXICAL RESOURCES

Traditional lexical resources (e.g., thesauri) can provide antonyms and synonyms for words and phrases already in the custom lexicon. For example, the Oxford English Dictionary online thesaurus lists an additional 6 synonyms for the noun "midnight" (Oxford Dictionaries, 2018). Recently more advanced resources have become available, which include full parts of speech tagging which relates words in a graph like structure known as a synset. WordNet (Princeton University, 2018) is a well-known example, which is supported in application programming interfaces such as the Python package NTLK (Bird et al., 2009).

Derivation and inflections are modifiers which are applied to words and include pluralization (e.g. suffixing a noun with "s") and changes to tense. Lexical software tools are available to provide derivations and inflections for a given word (WordNet exposes an API called morphy). Similar tools are available to provide common miss-spellings of words.

In the present study, pluralization was used but there was no automatic generation of derivations or inflections. Miss-spellings were included, but these were derived from researcher knowledge and vocabulary and derived from reviewing samples of opening posts.

NUMERICS

Whole and decimal numbers (e.g. 2, 3.5) and the numerical expression of times e.g. 1:15 can be parsed with regular expressions. This was the approach used and is documented in Appendix C.

RESEARCHER KNOWLEDGE AND VOCABULARY

Researchers will typically have a-priori knowledge of the field in which they are conducting research and a reasonable command of their native language. In this instance, all the authors engage in recreational fishing and this experience was used as the initial source from which the lexicon was compiled. Researcher knowledge was also used for common words, such as quantifiers (e.g. couple, one).

TARGETED REVIEW OF PROPER NOUNS

Proper nouns are not published in standard lexical sources and will not generally be known to researchers, experts or participants however, proper nouns could provide important information on the activity (e.g. a hunter may name a firearm manufacturer). If no suitable scholarly articles or reports are available which provide a definitive list then proper nouns can be compiled by using a targeted review of World Wide Web and other participant knowledge sources (e.g. magazines which carry manufacturer advertisements).

For-hire boat names were collated from the results of a Google search (*“charter boat” AND wales*), primarily from an online directory service. The classified sections of the June, July and August 2013 editions of the UK magazine “Sea Angler” (Sea Angler Magazine, 2018) were also reviewed for the boat names of operating skippers. The number of for-hire boats identified (n = 56) matched that identified by Richardson (2006). Kayak and private boat manufacturers and models were added from the primary author’s knowledge.

TEXT CONTENT REVIEW

Approximately 10% of scraped trips which contained a bass synonym (compiled from researcher knowledge and vocabulary and expert and participant vocabulary) were randomly selected. The text was reviewed manually and pertinent words and short added to the lexicon.

Table 8-1. How word types were compiled for entry into the custom lexicon.

Word Type	Used For	Source	Example
European seabass synonyms	Filter irrelevant content	Researcher Knowledge and Vocabulary; Expert and Participant Vocabulary	Schooly, bass, silver
Quantities: Quantifiers and determiners	Gear number, Catch number, Fish measures of length or weight	Researcher Knowledge and Vocabulary; Text Content Review	Quarter, 1/3, few, couple
Metrological nouns	Extract seabass measures (weight or length)	Researcher Knowledge and Vocabulary; Text Content Review	Kilo, kg, pound, lbs
Time (nouns)	Trip duration	Researcher Knowledge and Vocabulary; Text Content Review	½ an hour, half an hour, midnight
Platform related words	Platform, i.e. for-hire, private boat or shore.	Researcher Knowledge and Vocabulary, Targeted Review of Proper Nouns; Expert and Participant Vocabulary; Text Content Review	Charter, boat, skipper
Duration related words	Trip duration	Researcher Knowledge and Vocabulary; Text Content Review	Arrived, before, ended
Time and numerics	Catch number, seabass length and weight, gear number, trip duration	Regular Expression (See Supplementary Appendix B)	1.12, 3, 12:15

Equation Section (Next)

8.2 Chapter 4

Appendix E. List of Evaluated Sources

List of sea angling related sources evaluated for relevance. Suitability key as follows: R, source has a degree of relevance to the study but contains no suitable data; S, contains study relevant data; U, source found to have limited to no congruency with this report.

Source	Coverage	Date	Source type	compiler	owner	Description	Collection method(s)	Suitability

Study into Inland and Sea Fisheries in Wales	Wales	2000	Report	Nautilus Consultants	Welsh Government	Estimates based entirely on fisher knowledge of 2 people	Fisher and expert knowledge.	R – Contextual relevance
The economic value of recreational angling on the Dee estuary	Wales, North	2007	Thesis	Lee	Bangor University			R – Contextual relevance
Wales Activity Mapping: Economic evaluation of marine recreation activity. Business Survey.	Wales, Pembrokeshire	2012	Survey	Marine Planning Consultants	Welsh Government	Small self-selecting survey of businesses, combined with larger scale Wales Activity Mapping project over 2008-2010. Primary output is value per activity per location.	Self-selecting survey	R – Contextual relevance
Substance - Social and community benefits of angling	UK	1905	Survey	Substance	Substance	Extensive reports with a focus on socioeconomics. Results obfuscated by fresh and game angling. Data poor for Wales with no differentiation by platform.	Self-selecting questionnaire based methods.	R – Contextual relevance.
Value of [..] MPAs to [divers and anglers]	UK	2012	Survey	Multiple	Multiple	MPA centric, self-selecting online survey of anglers and divers, disseminated online, emails to club members and traditional advertising methods. No distinction made between charter and private boat. In depth WTP style economic analysis.	Various self-selection based instruments.	R – Contextual relevance.
The ecological impact of intertidal recreational bait collection	Unspecified	2004	Thesis	Harries	Bangor University	MSc thesis to read		R – Contextual relevance.
Fishing bait collection in the Menai Strait	Wales, North	1983	Thesis	P. Coates	P. Coates	Focuses on the exploitation of RSA bait species in the Menai Strait	Experimental, observational and expert knowledge	R – Contextual relevance.
The environmental impacts of bait-digging at Lleiniog Beach, Anglesey.	Wales, North	1994	Thesis	Spikes	Bangor University	Self-explanatory		R – Contextual relevance.
The tourism and recreational carrying capacity of Anglesey's coastal destinations Rhosneiger and Benllech	Wales, North	2006	Thesis	Hesketh	Bangor University	MSc thesis to read		R – Contextual relevance.
Socioeconomic [..] implications [..] to marine resource management for Wales	Wales	2006	Survey	E. A. Richardson	Bangor University	Comprehensive geographically specific work for doctorate thesis, Bangor University. Includes very extensive effort and economic surveys with excellent coverage of charter boat sector and economic analysis of the recreational sector.	Self-selection based instruments and face to face interviews over 2+ years.	S – Comprehensive with excellent coverage.
FishMap Môn	Wales, Anglesey	2012	Survey	Natural Resources Wales	Natural Resources Wales	None randomised creel survey primarily across Anglesey and approximately 50 miles of surrounding mainland coast.	None randomised creel.	S – Detailed spatial effort for all platforms

Wales Activity Mapping: CORE	Wales, South West	2008–2010	Survey	Multiple	Multiple	Creel, offsite, self-selected and expert instruments used in data collection. Does NOT appear to differentiate by platform, however locations are mapped, hence Boat/Shore can be inferred. Multiple organisations involved. NRW, PCF, EA, Crown Estate.	Expert knowledge and observation.	S – Detailed spatial effort for shore and boat
North Wales Pilot Surveys	Wales, North	2008	Survey	Marine Ecological Solutions	Natural Resources Wales	Onsite survey aimed primarily at effort and catch assessment of shore angling in North Wales. Only a single charter boat trip sampled for species catch composition. Makes recommendations for future survey approaches.	None randomised creel.	S – Detailed spatial effort for shore platform
Sea Angler Magazine Trophy Catches	Wales	1974–2003	Fisher knowledge	Richardson (2006)	N/A	RSAs submit catch records to Sea Angler magazine for Wales, transcribed direct from paper magazine.	RSA recorded	S – Limitations in use due to bias need to be considered.
Web and social media	Wales	Various	Fisher knowledge	N/A	N/A	Multiple sites exist, providing expert knowledge on fishing venues locations. Data aggregation services (e.g. google) can give proxies of popularity.	Manual review of sources.	S – Limitations in use due to bias need to be considered.
Angling Guide to Wales	Wales	1975	Grey literature	Clive Gammon	Unknown	Self-explanatory	Fisher knowledge	S – Locations
Shore Fishing: A Guide to Cardigan Bay	Wales, Mid	2013	Grey literature	John Mason	John Mason	Self-explanatory	Fisher knowledge	S – Locations
Lavers guide to sea angling in North Wales and Merseyside	Wales, North	1999	Grey literature	Phil Simpson	Lavers	Self-explanatory	Fisher knowledge	S – Locations
Sea Fishing North Wales and Anglesey	Wales, North	1968	Grey literature	Anthony Pearson	Publisher	Self-explanatory	Fisher knowledge	S – Locations
Sea Angling 2012 Internet and creel surveys	England	2012	Survey	CEFAS	CEFAS/MMO	First statistically rigorous sea angling survey in UK. Multiple instruments were used in economic, effort and catch assessments	Stratified random for effort with ONS standards compliant national survey. Self-selection elements in economic assessment.	S – Population & economic expansion though stratification transfer.
Opinions and Lifestyle Survey 2012	UK	2012	Organised Survey	Office of National Statistics	Office of National Statistics	Statistically sound national doorstep survey in which SA2012 questions were inserted. Includes respondents from Wales. Primarily used in population expansion for detailed economic, effort and catch estimates for SA2012, BUT, this has respondents from Wales.	Stratified random, ONS standards compliant.	S – Population & economic expansion though stratification transfer
UK Tourism Statistics (Multiple Years)	UK	2013	Survey	TNS Global	Multiple	National survey, held annually on both domestic and overseas visitors. Report does not contain angling data as participation is low, but evidence suggests it is collected (data referenced in other reports).	Stratified random, ONS style survey.	S – Population & economic expansion though stratification transfer.
Economic Impact of Outdoor Activity Tourism in Wales	Wales	2014	Survey	Miller Research	Visit Wales	Industry targeted interviews and self-selecting internet survey of participants to collect quantitative economic and activity participation level data. Participant targeting of visitors to Wales was primarily though social media channels.	Self-selection and expert knowledge.	S – Population & economic expansion though stratification transfer.

Great Britain Day Visitor Survey	UK	2011-2014	Survey	TNS Global	Visit England, Visit Scotland, Visit Wales	AS GBTS, but Day visits	Stratified random, ONS style survey.	S – Population & economic expansion though stratification transfer.
Great Britain Domestic Overnight Holidays	UK	2011–2014	Survey	TNS Global	Visit England, Visit Scotland, Visit Wales	AS GBTS, but holidays involving a minimum of 1 nights stay.	Stratified random, ONS style survey.	S – Population & economic expansion though stratification transfer.
Great Britain Tourism Survey	UK	2011–2014	Survey	TNS Global	Visit England, Visit Scotland, Visit Wales	UK wide, but has good coverage of wales (sponsored by Visit Wales). Unknown if addresses angler activity, but has spend figures for accommodation (for e.g.). Data source of the Great Britain Day Visitor Reports and the Domestic Overnight Tourism Reports.	Stratified random, ONS style survey.	S – Population & economic expansion though stratification transfer.
RSGB Omnibus Survey	UK	2003	Survey	Research Surveys of Great Britain (Taylor Nelson Sofres)	Unknown	Household survey, not angling specific. Used for population expansion in Drew (2004) study.	Stratified random, ONS style survey.	S – Provides historical comparative estimates through Drew (2004). Raw data unavailable.
Drew Associates	England & Wales	2003	Survey	Drew Associates	DEFRA	Comprehensive pseudo random creel and postal survey instruments, population expansion using RSGB Omnibus results. Postal survey from club membership frame. Postal and telephone survey of businesses for economic assessment.	Stratified random, ONS style survey.	S – Provides historical comparative estimates. Raw data unavailable.
CEFAS Small Boats Census.	Wales	To 2012	Survey	CEFAS	CEFAS/MMO	Coastal survey of small boats landing fish into ports including Wales with a recreational metier.	Intercept and observation survey	S – Though limited spatial resolution and coastal observation only.
Club Match Cards	Wales	Various	Fisher knowledge	N/A	RSA Entities	Record cards, recorded by anglers as part of organised or informal competitive angling.	RSA recorded	U – Bias and use permission problematic.
Ramsey Sound MCZ Assessment.	Wales		Survey	NRW	NRW	Data collected as part of MCZ assessments	Expert knowledge and observation.	U – Insufficient detail.
MMBFC Members Diaries	Wales, South	To 2012	Fisher knowledge	MMBFC	MMBFC	Mumbles Motor Boat fishing club catch diary data to 2012, centred on Swansea Bay area.	Fisher knowledge	U – Insufficient spatial extent.
Sea Fishing Atlas of Wales	Wales	To 2010	Fisher knowledge	NRW	NRW	GIS mapped metiers of commercial gears from expert knowledge sourced from WAG fisheries agencies and professional fisher bodies relating to 2000 - 2010. Broad scale maps on hobby netting.	Fisher and expert knowledge from fisheries experts and the fishing industry. Supplemented by information contained in published reports. Give a general indication of fishing activity over the period 2000 to 2005	U – No RSA data.

Recreational Bass Angling in Wales: [..] Data collection [and] Effort Distribution	Wales	2013	Thesis	G. Monkman	Bangor University	Assessment of angler diary records and effort distribution (primarily of bass anglers) across shore, private boat and charter records.	Social media data mining	U – Restricted to bass.
When the tide goes out - The biodiversity and conservation of the shores of Wales	Wales	1996–2006	Survey	CCW	Natural Resources Wales	10 year survey of intertidal habitats across wales, with extensive GIS mapping. Includes what may be an extensive source of angling effort distribution and bait collection impacts.	Observational and expert knowledge	U – Spatial scale too large, data sparse for RSA.
Time Use Survey	UK	Various	Survey	Unknown	Unknown	Stratified Random questionnaire based survey of activity. Need to examine further what data is available.	Stratified random, ONS style survey.	U – Sufficient detail unavailable.
Fishing Campaign Evaluation	Wales	2005	Survey	Beaufort Research	Beaufort Research	Survey aimed at assessment of visiting anglers, and their response to a fishing campaign. Raw data would be required to extract economic data, as no differentiation made between sea and fresh water anglers for spend figures.	Self-selection based survey from registered respondents.	U – Sufficient detail unavailable.
Bass Anglers Sports fishing Society	UK	Various	Fisher knowledge	BASS	RSA Entities	Catch cards and records from BASS, data poor for wales. Time series probably weak	RSA submitted	U – Too data sparse for Wales.
National Federation of Sea Anglers (NFSA) Specimen Records	UK	1976–2001	Fisher knowledge	NFSA	Unknown	Records of fish captures submitted and compiled by the now defunct NFSA	RSA submitted	U – Very poor spatial resolution.
Bass Fishing from the Shore in South East Wales	Wales, South	2009	Grey literature			Self-explanatory	Fisher knowledge	Un – Unavailable

Appendix F. Classification Summary of Candidate Sources

Table 8-2. Number of candidate sources identified as likely to contain spatially relevant data on marine recreational fishing activity across Wales. Used candidate sources had spatially reference data consumed by this study and unused candidate data were identified as containing no relevant data, or data which were unavailable (No access) under review. Private boat marine recreational fishing activity is included in used candidate sources however, this article did not include private boat activity in the analysis.

Suitable candidate sources

Data type	Study vehicle	Private boat only	Shore only	Shore and Private Boat
Field observation	GO or NGO report	1 (4%)		
Local knowledge	Angling literature		5 (20%)	
	www		14 (56%)	
Satellite imagery	www	1 (4%)		
Survey (Self-selecting or Non-randomised)	Academic thesis		1 (4%)	
	GO or NGO report		1 (4%)	2 (8%)

Unsuitable candidate sources

Data type	Study vehicle	Reason for exclusion				
		Multiple	No access	No use measure	Spatial resolution	Spatial scope
Field observation	Academic thesis					1 (3%)
	GO/NGO report	1 (3%)			1 (3%)	
Local knowledge	Diary or Logbook				3 (9%)	
	GO or NGO report				2 (6%)	
	Angling literature		2 (6%)			
	www			1 (3%)		
Review/Imputation from primary sources	GO or NGO report					2 (6%)
Survey (Randomised)	GO or NGO report	4 (12%)	1 (3%)		1 (3%)	2 (6%)
Survey (Self-selecting or Non-randomised)	Academic thesis		1 (3%)			
	GO or NGO	4 (12%)			1 (3%)	4 (12%)
	Non-randomised					1 (3%)

8.3 Chapter 5

Appendix G. lenscorrection.py

```
'''The calibrate function is a snippet of code from a larger
library.
calibrate will produce intrinsic radial and tangential distortion
matrices
for a given camera and focal length.
This profile can then be used to undistort images captured with that
camera
model and focal length.

It produces calibration profiles for the thin lens distortion model
and the thin lens model using the OpenCV API for Python.

https://docs.opencv.org/3.4/db/d58/group__calib3d__fisheye.html
https://docs.opencv.org/3.4.1/dc/dbb/tutorial_py_calibration.html
'''
#NOTE:
#This is not a standalone script, but contains the core functions
required to
#create fisheye and thin lens camera profiles. However, it would be
necessary
#to write the code to save and retrieve camera profiles (see inline
comments)



#Python base modules
import os as _os
import glob as _glob
import pickle as _pickle
import itertools as _itertools

#Third party modules which require installation using a package
manager (e.g. pip)
import imghdr as _imghdr
import cv2 as _cv2
import numpy as _np


FISHEYE_CALIBRATION_FLAGS = _cv2.fisheye.CALIB_RECOMPUTE_EXTRINSIC +
_cv2.fisheye.CALIB_CHECK_COND + _cv2.fisheye.CALIB_FIX_SKEW
```

```
TERMINATION_CRITERIA_FISHEYE = (_cv2.TERM_CRITERIA_COUNT, 50, 1e-6)
TERINATION_CRITERIA_SUBPIX = (_cv2.TERM_CRITERIA_EPS +
_cv2.TERM_CRITERIA_COUNT, 30, 0.1)
MESSAGES = []


def calibrate(self, image_os.paths, width, height,
skip_fisheye=False, fisheye_no_check=True):
   '''(list, float, float, bool, bool) -> list, list
   Calibrate camera from images

   image_os.paths:
      a list of paths where images are saved
   width, height:
      use images of this resolution in image_os.paths
   height
   skip_fisheye:
      self explanatory
   fisheye_no_check:
      stop CV doing a checks

   returns:
      list of bad images, list of deleted images
   '''
   obj_points = []
   img_points = []
   img_points_fisheye = []
   fcnt = 0
   cnt = 0
   bad_images = []
   deleted_images = []

   FE_CALIB_FLAGS = FISHEYE_CALIBRATION_FLAGS
   if fisheye_no_check:
      FE_CALIB_FLAGS -= _cv2.fisheye.CALIB_CHECK_COND
   image_os.paths = [x for x in
_iolib.file_list_glob_generator(image_os.paths)]
   image_os.paths_ok = []
   for fn in image_os.paths:
      if _info.ImageInfo.is_image(fn):
         img = _cv2.imread(_os.path.normpath(fn), 0)
         w, h = img.shape[1], img.shape[0]
```

```
        if w == width and h == height:
            cnt += 1
            found, corners = _cv2.findChessboardCorners(
                img, self.pattern_size,
flags=_cv2.CALIB_CB_ADAPTIVE_THRESH + _cv2.CALIB_CB_NORMALIZE_IMAGE)
            if found:
                fcnt += 1
                _cv2.cornerSubPix(img, corners, (10, 10), (-1, -1),
Calibration.TERINATION_CRITERIA_SUBPIX)
                img_points.append(corners.reshape(-1, 2))
                img_points_fisheye.append(corners.reshape(1, -1, 2))
                obj_points.append(self._pattern_points)
                image_os.paths_ok.append(fn)
            else:
                MESSAGES.append('Chessboard vertices not found in %s.
The file was deleted.' % (fn))
                try:
                    _os.remove(fn)
                    print(MESSAGES[-1])
                except Exception as _
                    pass

    if not img_points:
        raise ValueError('Failed to find any vertices in any images.
OpenCV findChessboardCorners is bugged, pattern size must be 9 x 6
vertices in photo and ini file.')

    n_ok = len(img_points_fisheye)

    # calculate camera distortion
    rms, camera_matrix, dist_coefs, rvecs, tvecs =
_cv2.calibrateCamera(
        obj_points, img_points, (width, height), None, None)

    cm = _pickle.dumps(camera_matrix, _pickle.HIGHEST_PROTOCOL)
    dc = _pickle.dumps(dist_coefs, _pickle.HIGHEST_PROTOCOL)
    rv = _pickle.dumps(rvecs, _pickle.HIGHEST_PROTOCOL)
    tv = _pickle.dumps(tvecs, _pickle.HIGHEST_PROTOCOL)

    if not skip_fisheye:
        #K and D passed by ref in fisheye.calibrate. Initialise them
first.
```

```
      K = _np.zeros((3, 3))
      D = _np.zeros((4, 1))

      #pattern_points is a tuple with the number of x and y vertices
of the chess board
      #ie (9,6) would be a chessboard with 9 x 6 vertices
      chessboard_model = _np.zeros((1, self.pattern_size[0] *
self.pattern_size[1], 3), dtype=_np.float32)
      chessboard_model[0, :, :2] = _np.mgrid[0:self.pattern_size[0],
0:self.pattern_size[1]].T.reshape(-1, 2)

      #this delete invalid images and deletes the detected points
and chessboard model from
      #the numpy array,
      bad_images = []
      deleted_images = []
      while _np.array_equal(K, _np.zeros((3, 3))):
         try:
            rvecs = [_np.zeros((1, 1, 3), dtype=_np.float64) for i in
range(n_ok)]
            tvecs = [_np.zeros((1, 1, 3), dtype=_np.float64) for i in
range(n_ok)]
            #pattern_points is a tuple with the number of x and y
vertices of the chess board
            #ie (9,6) would be a chessboard with 9 x 6 vertices
            rms, _, _, _, _ = _cv2.fisheye.calibrate(
               [chessboard_model]*n_ok, img_points_fisheye, (width,
height),
               K, D, rvecs, tvecs, FE_CALIB_FLAGS,
Calibration.TERMINATION_CRITERIA_FISHEYE)
         except _cv2.error as e: #we need to parse the index out of
the error message
            ind = int(get_between(str(e), 'input array ', ' in
function'))
            assert isinstance(ind, int) and ind >= 0, 'Could not
determine the index of the bad calibration image, perhaps the
exception text has changed'
            bad_images.append(image_os.paths_ok[ind])
            n = [x for x in range(0, ind)]
            _ = ([n.append(y) for y in range(ind+1, n_ok)])
            n_ok -= 1
            lst = _np.array(img_points_fisheye)
            lst = _np.squeeze(lst[[n], ...], 0)
```

```
            img_points_fisheye = [x  for x in lst] #rebuild as list
of numpy arrays

            try:
               _os.remove(image_os.paths_ok[ind])
               if not file_exists(image_os.paths_ok[ind]):
                  deleted_images.append(image_os.paths_ok[ind])
            except Exception as _
               pass

            del image_os.paths_ok[ind]
            assert len(image_os.paths_ok) == len(img_points_fisheye),
'Number of remaining valid calibration images in list
image_os.paths_ok does not match the number of images which have
points in img_points_fisheye'

      kk = _pickle.dumps(K, _pickle.HIGHEST_PROTOCOL)
      dd = _pickle.dumps(D, _pickle.HIGHEST_PROTOCOL)

   else:
      kk = None
      dd = None

   #The variables cm, dc, rms, rv, tv, kk, dd
   #now contain all the values relating to the intrinsic property
   #of the lens, these should be exported to persistent storage.

   #The following code WILL NOT WORK, as it relies on a custom
library
   #created by the primary author.
   #The code persists the variables to an sqlite database.
   #Adapt to own requirements. The arrays could just be dumped
   #to the file system with suitable name and reloaded later
   with _lenscorrectiondb.Conn(cnstr=_CALIBRATION_CONNECTION_STRING)
as conn:
      db = _lenscorrectiondb.CalibrationCRUD(conn)
      modelid = int(db.crud_camera_upsert(self.camera_model))

      #calibrationid returned by crud_calibration_upsert, but we
dont need it
      _ = int(db.crud_calibration_upsert(
         modelid, width, height, cm, dc, rms, rv, tv, kk, dd))
```

```
      conn.commit()
   #END OF NON FUNCTIONAL CODE
   return bad_images, deleted_images


def undistort(
         cam,
         imgpaths_or_imagelist,
         outpath,
         label='_UND',
         label_fisheye='_FISHUND',
         crop=True,
         use_fisheye=False,
      cma=None, dcoef=None, rvect=None, tvect=None, K=None, D=None):
    '''(Camera, str|iterable, str, str, bool) -> void
    Bulk undistort, reading in the camera profile according to model
name as matched in lenscorrection.py.ini
    Multiple paths can be provided

    imgpaths_or_imagelist can be an iterable of paths or a list. If
appears to be paths,
    then _glob will be combined with known image extensions to list
all files in paths
    which appear to be images. If a single directory string is
passed in, this
    will also be valid and globbed.

    Converted images are saved to outpath, with label appended to
the original file name.

   IMPORTANT NOTE:
   cma, dcoef, rvect, tvect, K and D are the parameters created from
   the calibrate function, and would need to be loaded (e.g. from
pickled variables)
   before calling undistort. The code to load is not included in
this document
   as the primary author saved these values to an sqlite database.
    '''
    blobs = {}
    useglob = True
    bad_res = []
    subst_res = []
```

```
    if isinstance(imgpaths_or_imagelist, str):
        imgpaths_or_imagelist = [imgpaths_or_imagelist]
    else:
        # look to see if the list is mostly (50%) valid files rather
than
        # directories
        validcnt = 0.0
        for myfiles in imgpaths_or_imagelist:
            validcnt += _os.path.isfile(_os.path.normpath(myfiles))
        if validcnt / len(imgpaths_or_imagelist) > 0.5:
            useglob = False

    if useglob:
        globlist = file_list_generator(
            imgpaths_or_imagelist,
            ['jpg', 'jpeg', 'png', 'gif'])
        newlist = []
        for wildcards in globlist:
            for fil in _glob(wildcards):
                newlist.append(fil)
    else:
        newlist = imgpaths_or_imagelist

    cnt = 1
    success = 0
    outpath = _os.path.normpath(outpath)
    create_folder(outpath)



    print('Undistort mode: %s' % ('fisheye lens model' if
use_fisheye else 'standard lens model'))

   last_width = 0
   last_height = 0
   for fil in newlist:
      try:
         resize_suffix = ''
         # used later to rebuild output file name
         _, name, ext = get_file_parts(fil)
         orig_img = _cv2.imread(fil)
```

```
        width, height = orig_img.shape[1], orig_img.shape[0]
        if (last_width != width and last_height !=
                height) and height > 0 and width > 0:
            blobs = db.crud_read_calibration_blobs(
                cam.model, height, width)

        if blobs is None:
            s = 'No calibration data for image %s, resolution
[%sx%s]' % (fil, width, height)
            print(s)
        else:
            blobs = {'cmat':cmat, 'dcoef':dcoef, 'rvect':rvect,
'tvect':tvect, 'K':K, 'D':D}
            img = _undistort(cam, orig_img, blobs, crop,
use_fisheye=use_fisheye)
            if img is None:
                print('File %s failed in _undistort.\n' % (fil))
            else:
                if use_fisheye:
                    outfile = _os.path.join(outpath, name +
label_fisheye + resize_suffix + ext)
                else:
                    outfile = _os.path.join(outpath, name + label +
resize_suffix + ext)
                outfile = _os.path.normpath(outfile)
                _cv2.imwrite(outfile, img)

        last_width = width
        last_height = height

    except Exception as _:
        pass

    if bad_res:
        print(
            'Resolutions with no calibration matricies: %s' %
            (" ".join(bad_res)))



def _undistort(img, mats, crop=True, use_fisheye=True):
    '''[c]Camera, ndarray (image), dic, bool -> ndarray (image) |
None
```

```
    Undistorts an image based on the lens profile loaded into the
Camera class cam.
    dic is a dictionary containing the undistortion matrices
    {'cmat':cmat, 'dcoef':dcoef, 'rvect':rvect, 'tvect':tvect}


    Returns None if an exception occurs
    '''
    assert isinstance(img, _np.ndarray)
    try:
        h, w = img.shape[:2]
        if use_fisheye:
            R = _np.eye(3)
            #K stores just the focal length camera parameters and
the image centre
            #See
https://docs.opencv.org/2.4/doc/tutorials/calib3d/camera_calibration
/camera_calibration.html
            map1, map2 =
_cv2.fisheye.initUndistortRectifyMap(mats['K'], mats['D'], R,
mats['K'], (w, h), _cv2.CV_16SC2)
            dst = _cv2.remap(img, map1, map2,
interpolation=_cv2.INTER_LINEAR, borderMode=_cv2.BORDER_CONSTANT)
        else:
            newcameramtx, roi = _cv2.getOptimalNewCameraMatrix(
                mats['cmat'], mats['dcoef'], (w, h), 1, (w, h))
            dst = _cv2.undistort(
                img,
                mats['cmat'],
                mats['dcoef'],
                None,
                newcameramtx)
            if roi == (0, 0, 0, 0):
                _warn.warn('_cv2.getOptimalNewCameraMatrix could not
identify the ROI. Try recalibrating with more small calibration
images at the camera edge or sets of larger calibration
images.\n\nImages were undistorted but should be checked.')
            else:
                if crop:
                    x, y, w, h = roi
                    dst = dst[y:y + h, x:x + w]


    except Exception:
        print(Exception.message)
```

```
        dst = None
    finally:
        return dst



#Helper functions
def file_list_glob_generator(wilded_os.path, recurse=False):
    '''(str, bool)->yields strings (file paths)
    _glob.glob generator from wildcarded path
    Wilded path would be something like 'c:/*.tmp' or c:/*.*

    Yields actual file names, e.g. c:/temp/a.tmp

    SUPPORTS RECURSION
    '''
    fld, f = get_file_parts2(wilded_os.path)[0:2]

    if recurse:
        wilded_os.path = _os.path.normpath(_os.path.join(fld, '**',
f))

    for file in _glob.iglob(wilded_os.path, recursive=recurse):
        yield _os.path.normpath(file)


def is_image(file_os.path, try_load=False):
   '''(str, bool) ->b ool
   Pass in a file string and see if it looks like an image.

   try_load:
      Try and loading file_os.path with _cv2.imread (costly)

   Example:
      >>>is_image('C:/temp/picture.jpg', try_load=False)
      True
   '''
   ret = False
   file_os.path = _os.path.normpath(file_os.path)
```

```
    if not _os.path.isfile(file_os.path):
        return False

    if not _imghdr.what(file_os.path) is None:
        ret = True
        if try_load:
            try:
                ret = False
                img = _cv2.imread(file_os.path)
                if isinstance(img, _np.ndarray):
                    ret = True
            except Exception as dummy:
                pass

    return ret


def file_list_generator(paths, wildcards):
    '''(iterable, iterable) -> tuple
    Takes a list of paths and wildcards and creates a
    generator which can be used to iterate through
    the generated file list so:
    paths = ('c:/','d:/')     wildcards=('*.ini','*.txt')
    Will generate: c:/*.ini, c:/*.txt, d:/*.ini, d:/*.txt

    ie. Yields wildcards for consumption a _glob.glob.
    '''
    if isinstance(wildcards, str):
        wildcards = [wildcards]

    ww = ['*' + x if x[0] == '.' else x for x in wildcards]

    for vals in (_stringslib.add_right(x[0]) + x[1]
                 for x in _itertools.product(paths, ww)):
        yield _os.path.normpath(vals)


def get_between(s, first, last):
    '''(str, str, str) -> str
    Gets text between first and last, searching from the left
```

```
    s:
        String to search
    first:
        first substring
    last:
        last substring
    '''
    try:
        start = s.index( first ) + len( first )
        end = s.index( last, start )
        return s[start:end]
    except ValueError:
        return ''


def file_exists(file_name):
    '''(str) -> bool
    Returns true if file exists
    '''
    if isinstance(file_name, str):
        return _os.path.isfile(fixp(file_name))
    return False


def create_folder(folder_name):
    '''(str) -> void
    creates a folder
    '''
    if not _os.path.exists(folder_name):
        _os.makedirs(folder_name)


def get_file_parts(filepath):
    '''(str)->list[path, filepart, extension]
    Given path to a file, split it into path,
    file part and extension.

    filepath:
        full path to a file.
```

```
Returns:
    The folder, the filename without the extension
    and the extension

Example:
>>>get_file_parts('c:/temp/myfile.txt')
'c:/temp', 'myfile', '.txt'
'''
folder, fname = _os.path.split(filepath)
fname, ext = _os.path.splitext(fname)
return [folder, fname, ext]
```

Appendix H. shape_area.py

```
# pylint: disable=C0103, too-few-public-methods, locally-disabled,
no-self-use, unused-argument, unused-import
'''find height at 50% of shape area, prints it to screen.
Used on an cross sectional image of a fish, where
the fish is posed against a white background and
the fish profile is darker than the background.

Height is a proportion of the total height.
'''

import argparse



import cv2
import numpy as np




def main():
    '''main'''
    cmdline = argparse.ArgumentParser(description=__doc__)
    cmdline.add_argument('-rt', help='Reverse and threshhold the
image', action='store_true')
    cmdline.add_argument('imgfile', help='File with the image')
    args = cmdline.parse_args()

    img = cv2.imread(args.imgfile, 0)
    img = cv2.cvtColor(img, cv2.COLOR_BGR2GRAY)
    if args.rt:
        #correction for some dark grey artifacts in image after
        #binarizing the image
        img[img < 50] = 51
        img[img > 51] = 0
        img[img == 51] = 255
        #now the fish profile is white on black

    assert isinstance(img, np.ndarray)

    n = 0
```

```
    pxcnt = 0
    pxmax = 0
    for _, col in enumerate(img.T):
        if 255 in col:
            n += 1
            pxcnt += len(col[col == 255])
            pxmax = len(col[col == 255]) if len(col[col == 255]) >
pxmax else pxmax

    print('Standard mean height: %.3f' % ((pxcnt/n)/pxmax))



if __name__ == '__main__':
    main()
```

Appendix I. perspective.py

```python
'''Corrections for the width profile of an object in single camera photogrammetry'''

class Camera():
    '''just a container for camera properties
    '''
    def __init__(self, f, px_x, px_y, x_mm, y_mm):
        self.f = f
        self.px_x = px_x
        self.px_y = px_y
        self.x_mm = x_mm
        self.y_mm = y_mm


class Measure():
    '''a measure, just a variable container'''
    def __init__(self, lens_subj_dist=None, marker_length_mm=None, marker_length_px=None):
        self.lens_subj_dist = lens_subj_dist
        self.marker_length_mm = marker_length_mm
        self.marker_length_px = marker_length_px


def get_perspective_correction(bg_dist, object_depth, length):
    '''(float, float)->float|None
    Return the length corrected for the depth of the object
    considering the backplane of the object to be the best
    representative of the length
    *NOTE* The length of the object has been accurately measured
    '''
    if bg_dist is None or object_depth is None or length is None:
        return None
    elif bg_dist == 0 or 1 - (object_depth / bg_dist) == 0:
        return None

    return length / (1 - (object_depth / bg_dist))
```

```


def get_perspective_correction_iter_linear(coeff,
                                           const,
                                           bg_dist,
                                           length,
                                           profile_mean_height=1,
                                           last_length=0,
                                           stop_below_proportion=0.01):
    '''(float, float, float, float,float)->float|None
    Return the length corrected for the depth of the object
    considering the backplane of the object to be the best
    representative of the length.
    *NOTE* The length of the object was itself estimated from the foreground standard measure

     Coeff and constant are used to calculate an objects depth from its length
     The object depth is used to create an iterative series sum which adds to the length
     to return the sum of lengths once the last length added was less then stop_below

    stop_below_proportion is the stopping criteria, once the last
            calculated length to add is is less than last_length*stop_below_proportion
    we return the result and stop the iteration
    '''
    if bg_dist == 0 or bg_dist is None or coeff == 0 or coeff is None or length is None:
        return None

    if last_length == 0:
        object_depth = length * coeff + const
    else:
        object_depth = last_length * coeff + const

```

```
    if object_depth <= 0:
        return length
    elif length == 0:
        return 0
    elif (last_length / length < stop_below_proportion) and last_length > 0:
        return length

    if last_length == 0:  # first call
        l = get_perspective_correction(bg_dist, object_depth, length) - length
    else:
        l = get_perspective_correction(bg_dist, object_depth, last_length) - last_length

    if l is None:
        return None

    return get_perspective_correction_iter_linear(coeff, const, bg_dist, length + (l * profile_mean_height), (l * profile_mean_height), stop_below_proportion)


def subjdist_knowndist(Known, Unknown):
    '''(Class:Measure, Class:Measure) -> float|None
    Get subject-lens distance
    estimate from a photograph of known distance
    with fiducial marker of known length
    '''
    assert isinstance(Known, Measure)
    assert isinstance(Unknown, Measure)
    x = [Known.marker_length_px, Known.lens_subj_dist, Known.marker_length_mm, Unknown.marker_length_mm, Unknown.marker_length_px]
    if not all(x):
        return None
```

```
    if Known.marker_length_mm == 0 or Unknown.marker_length_px == 0:
        return None
    F = Known.marker_length_px * Known.lens_subj_dist / Known.marker_length_mm
    return Unknown.marker_length_mm * F / Unknown.marker_length_px


def subjdist_camera(Cam, Unknown):
    '''(Class:Camera, Class:Measure) -> float|None

    Estimate lens-subject distance from the camera properties
    and the known marker length in mm and measure marker pixel
    length

    Currently assumes just using the width and not the height.

    Camera properties needed are the:
    Real cmos width in mm
    The cmos width in pixels
    The cameras focal length in mm
    '''
    assert isinstance(Cam, Camera)
    assert isinstance(Unknown, Measure)
    x = [Cam.f, Unknown.marker_length_mm, Cam.px_x, Unknown.marker_length_px, Cam.x_mm]
    if not all(x):
        return None

    return (Cam.f * Unknown.marker_length_mm * Cam.px_x) / (Unknown.marker_length_px * Cam.x_mm)
```

8.4 CHAPTER 6

Appendix J. Additional Methods

TRAINING AND VALIDATION IMAGE ACQUISITION

In computer vision, object detection identifies the location of a subject within an image. The most familiar example is the detection of human faces in images and during video capture. The computer is "shown" many examples of the subject (training images) from which the features (e.g. shape and patch colour) associated with the subject are learnt. Separate validation images are then used to determine how effective the learned model is at object localization. Usually training and validation images are partitioned by some ratio (e.g. 8:2) from the set of all available images.

Candidate training and validation images (*trainval images*) were acquired from Flikr and a Google image search using the Scrapy Python package (Scrapy, 2017). The Google search term was *bass (angling OR fishing) site:co.uk*. Images with a pose conducive to accurate length estimation (canonical view) were retained. Images with significant body occlusion, body contortion, perspective distortion or blurring were rejected (e.g. Figure 8-3).

Figure 8-3. Example bass images. Left, a rejected image, displaying body contortion and occlusion of the tail. Extracting an accurate length estimation by any means would be problematic. Right, a well-posed canonical image with snout, fork and fin tips clearly visible and no body contortion. The right image was taken with a GoPro Hero 5, note the radial distortion.

The appearance of the anal, pelvic, spiny and soft dorsal fins is highly variable between images. Hence the region of interest (RoI, a quadrilateral) for each image was drawn tight to the body, to the limits of the caudal fin tip(s) and the snout vertex, bounded by the body depth (Figure 6-1).

Figure 8-4. Examples of the defined region of interest (RoI) for training and validation images. Top, both tips of the caudal fin included in the RoI set tight to the lower and upper bounds of the body depth. Bottom, single caudal fin tip included. Images where no caudal fin tip would fall within the tight body bounding box were rejected.

ROIs were marked in VGG Image Annotator (Dutta, Gupta, & Zisserman, 2018). During pipeline processing, the bounding rectangle of the RoI was determined and the image and the bounding rectangle vertices rotated so the long edge were parallel with the x-axis. Negative training and validation images were implicitly sampled from areas of the positive images outside of the defined ROIs. Images, image metadata and class labels were exported to Google’s TFRecord format. A total of 918 canonical images were selected from scraped images. 734 (80%) were used for the training set and 184 (20%) for the model validation set. In addition, 30 negative images were created from regions of the marker images outside of the RoI to test trained CNNs for false positive detections.

Images were managed in digiKam (DigiKam, 2018). DigiKam image metadata (e.g. user tags and file system location) are stored in an SQLite database (Hipp, Kennedy, & Mistachkin, 2018) which can be accessed programmatically and combined within image processing pipelines e.g. to exclude images based on custom tags.

TENSORFLOW

Tensorflow (TF) is an open source machine learning API (Abadi et al., 2015), with extensive support for convolutional neural networks (CNN). TF is authored by Google and available for researchers to use in Python or C++. TF supports NVIDIA’s CUDA architecture (NVIDIA, 2018) to efficiently run deep learning networks on NVIDIA graphics cards however, training a deep CNN remains computationally expensive even on modern CUDA-enabled graphics cards. Tensorflow provides multiple pretrained deep CNN models in their object detection model zoo (Google, 2018), The selected models were pretrained by Google researchers on the COCO image set (COCO, 2018). The pretrained networks were (i) ResNet-101 (He et al., 2016), (ii) Single shot MobileNet detector (A. G. Howard et al., 2017) and (iii) NASNet (Zoph & Le, 2017); abbrevs. ResNet, MobileNet and NASNet respectively. These pretrained “frozen inference graphs” can be used for object recognition and detection by training the final fully connected class prediction layers of the model on new object classes. The compiled TFRecords were used to train the prediction layers for the three COCO (COCO, 2018) trained models. The three networks were chosen as they offer increasing speed performance at the cost of accuracy (Table 8-3). The single shot MobileNet detector is of particular interest because it can be deployed on mobile devices.

Table 8-3. Performance of the 3 pretrained R-CNNs NASNet (Google, 2018; Zoph & Le, 2017), ResNet-101 (He et al., 2016) and single shot MobileNet detector (A. G. Howard et al., 2017). COCO mAP is the mean average precision of the CNN when used to detect objects in the COCO image set (COCO, 2018). mAP is a measure of the number of successful class (e.g. dog, cat) identifications in multi-class detection problems.

Model name	**Speed (ms)**	**COCO mAP**
faster_rcnn_nas	1833	43
faster_rcnn_resnet101_coco	106	32
ssdlite_mobilenet_v1_coco	30	21

COMPUTE PLATFORM

All training and inference used Tensorflow 1.8 on 64-bit Python 3.5.4 with NVIDIA CUDA 9.0 and cuDNN 7.1.2, on an Intel Xeon-1650 CPU with 32 GB EEC RAM and an NVIDIA GTX 1060 6 GB graphics card with solid state drive. Additional Python packages were NumPy 1.14.2, OpenCV 3.4.0, SciPy 1.0.1, statsmodels 0.8.0 and Sympy 1.1.1.

INTERSECTION OVER UNION

This article uses intersection over union (IoU%) as an indicator of object detection accuracy (mean average precision is not applicable). For a single detection $IoU\% = 100 \cdot A_{gt} \cap A_{det} / A_{gt} \cup A_{det}$, where A_{gt} is the ground truth RoI area and A_{det} is the detection RoI area (Figure 8-5).

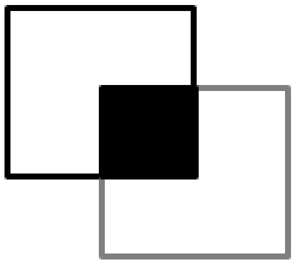

Figure 8-5, Intersection over union is the sum of the intersection area, divided by the sum of the intersection area and non-intersecting area. Giving a value between 0 and 1.

FIDUCIAL MARKER SELECTION AND IMAGE ACQUISITION

Fiducial markers (henceforth *markers*) need to be convenient to carry, and detection must be robust to perspective changes and image problems (e.g. overexposure and blur). ArUco markers (Garrido-Jurado et al., 2014) have detection support in OpenCV (OpenCV team, 2018) and preliminary assessment showed detection was robust for the use case under consideration. Three ArUco markers of side lengths 25mm, 30mm and 50mm were printed on matt vinyl and mounted on polypropylene sheets to produce markers similar to a credit card. Marker sizes were validated with an EC class 1 rule. Events where sea bass were captured from the shore and afloat (charter boats and a commercial longlining vessel) were attended and photographs of sea bass captured (Table 8-4).

Table 8-4. Count of sea bass images captured with a fiducial marker.

Camera\Platform	Shore	Afloat
Fujitsu XP30	63	45
GoPro Hero 5	61	50
S5690	58	49

The camera models were a GoPro Hero 5 (henceforth Hero 5), a Samsung S5690 smartphone (henceforth S5690) and a Fujifilm FinePix XP30 (henceforth XP30). These cameras were chosen as they are cheap, non-specialist cameras with differing amounts of radial distortion, and are also waterproof. All images were taken without mechanical zoom as different focal lengths require a separate profile for accurate correction. Fish were posed to minimise body distortion and occlusion, and approximately centred in the camera's field of view and filled 50% to 80% of the horizontal field of view. Total length (TL) is the length between the point at tip of the snout (snout vertex) and the midpoint of the line joining the tips of the caudal fin at natural minimal splay (caudal vertex). Reasonable measures were taken to reduce caudal fin splay according to standard practice when recording TL. The marker was placed as illustrated in Figure 6-3.

This placement puts the marker at the surface of the fish closest to the camera and the subject. Attention was given to minimising tangential distortion, but no special measurements or devices were employed.

Fish TL was recorded to the nearest millimetre and the length written on card and placed within the camera's field of view during image capture. If the pose of the fish changed, the process was restarted. The attended fishing events occurred in North and South Wales, UK (Figure 8-6).

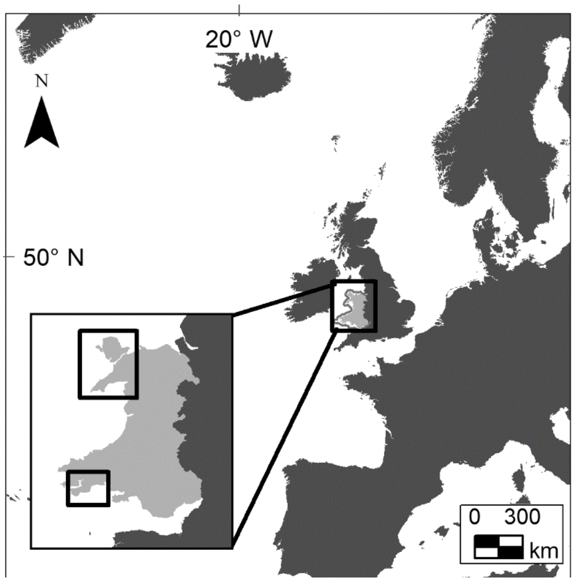

Figure 8-6. Study area.

Black boxes in inset show approximate area in which sea bass captures were made.

UNDISTORTING MARKER IMAGES

Object detection and classification should be relatively robust to image distortion if suitably trained however, subject distortion arising from the intrinsic properties of the camera lens require correction to reduce length estimation error. Correction is particularly important where radial distortion is large. It is evident that in estimating length, the world units per pixel across the extent of the subject within the photograph should be constant. The intrinsic parameters and the radial and tangential lens distortion at the default focal length were calculated using the OpenCV API (OpenCV team, 2018) and stored in an SQLite database (henceforth *profiles*). The Hero 5 supports 3 field of view settings (FoV) hence a separate profile was created for each FoV. S5690 and XP30 images were corrected for intrinsic radial and tangential distortion (henceforth *undistorted images*) using the standard camera calibration model (Z. Zhang, 2000). The profiles for the Hero 5 were created using a fisheye model (Scaramuzza et al., 2006) to achieve satisfactory performance because of the camera's marked radial distortion. Original images were undistorted using OpenCV to apply the corresponding camera profile. Images were resized (retaining their aspect ratio) using bilinear interpolation to reduce memory demands during processing. Images captured with the S5690 and XP30 cameras were resized using bilinear interpolation to 1280 x 960 pixels. Hero5 images were resized to 1440 x 1080 pixels. ROIs were marked in VGG Image Annotator (Dutta et al., 2018).

LENGTH ESTIMATION

LENS-SUBJECT DISTANCE ESTIMATION

To correct TL estimates of subjects in single camera photogrammetry with a foreground fiducial marker it is necessary to account for the changing distance between the lens, the fish and the fiducial marker (parallax error). If this is not accounted for, then TL will be underestimated. However, to apply an effective correction, the distance between the lens and the marker is required on a per image basis. Where data is to be collected under conditions over which the researcher has limited control (e.g. volunteer lead data collection) then this distance is impractical to measure. Hence, the lens-subject distance (*d*) must be estimated. This can be done by taking one (or more) calibration images with a fiducial marker of known length, according to the well-known relationship$d = (h_c \cdot h \cdot f)/(\acute{h_c} \cdot \acute{h})$, where f is the focal length, h_c is the real-world size of the calibration marker, h is the real-world size of the fiducial marker, $\acute{h_c}$ is the height in pixels of the calibration marker and $\acute{h}$ is the height of the fiducial marker in pixels.

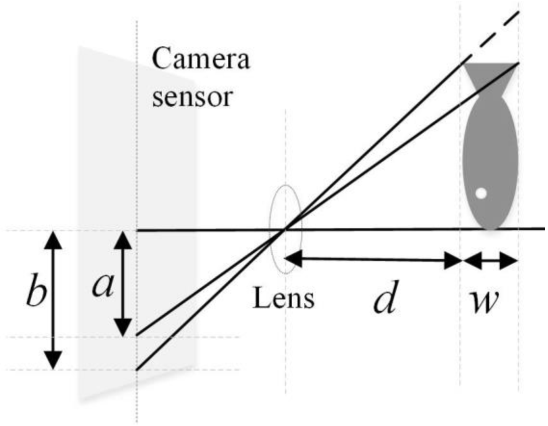

Figure 8-7. The thin-lens camera model.

Relates real-world lengths to image formation at the camera sensor. Reproduced from Monkman et al. (2018).

TOTAL LENGTH CORRECTION

This is the methodology of (Monkman, Hyder, Kaiser, & Vidal, 2018), which reduced mean bias error for fusiform and flat fish length estimation to < 2% when using foreground or parallel-laser fiducial markers when estimating fish length with cameras with significant radial distortion and unknown lens-subject distance. The true real world length per pixel represented by the fiducial marker would be measured at a plane parallel to the camera sensor which falls between *d* and $d + w$ in Figure 8-7.The lens-subject distance of this notional plane will be *d* plus the mean width of the fish profile. The mean width profile was calculated by dividing the fish in two through the axis of the coronal plane, photographing the fish and then thresholding the image. Fish were taken from frozen samples held by The University. Non-invasive methods (e.g. 3D model recreation) could also be used. OpenCV was used to calculate the mean pixel height (mph) of the width profile in the threshholded image. For sea bass, mph = 0.598 (Monkman, Hyder, et al., 2018). A further difficulty is that the total width of the fish is unknown. Fish width was calculated from $width = 0.136 \cdot total\ length\ - 0.367$ (Poli et al., 2001), using the initial TL estimate derived from the image.

LENGTH ESTIMATION UNDER ROTATION MODELS

OpenCV was used to produce in-memory images rotated from -30° to30° degrees in 1° degree increments (*theta*, θ). Detections under rotation were reviewed by the authors and Figure 8-8 was proposed as a representative model of the relationship between an ideal prediction of the subject under no rotation, and the bounding box detection under rotation. Based on the geometric relationship shown in Figure 8-8 the unknown side lengths a, b of the idealized detection are related to the detections under rotation side lengths A, B by the following 3 equations.

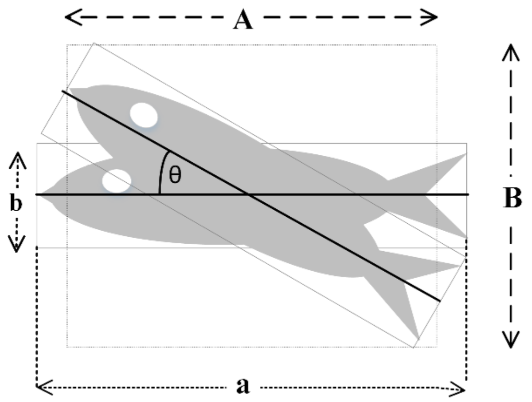

Figure 8-8. Rotation of subject through θ degrees. The lines a and b represent the (unknown) height and width of an ideal detection of an unrotated subject. The lengths A and B is an approximation made from observation of the side lengths of a detection bounding box of a subject rotated through θ degrees. The ratio a/b is known. This representation overestimated B, hence a correction was calculated empirically.

(7)
$$\frac{\boldsymbol{a}}{\boldsymbol{b}} = \boldsymbol{C}$$

(8)
$$\boldsymbol{A} = \boldsymbol{a}\cos(\theta)$$

(9)
$$\boldsymbol{B} = \boldsymbol{b}\cos(\theta) + \boldsymbol{a}\sin(\theta)$$

Where Eq. 7 is the mean length to depth ratio of the subject at $\theta = 0$. The constant C was calculated from the manually defined RoIs at $\theta = 0$ (C, mean ±S.D., 4.32 ±0.23). In Eq. 8 and Eq. 9, A and B are known for each image and can be reduced by substitution of Eq. 7 into Eq. 9 to two unknowns θ and b. The Python SciPy library (E. Jones et al., 2017) was used to find the roots of Eq. 8 and Eq. 9 after substitution.

DETECTION AND LENGTH ESTIMATION UNDER ROTATION, FLIPPING AND DOWNSAMPLING

R-CNNs are invariant to object translation, scaling and mild distortion however, they are not inherently rotation invariant. R-CNNs are capable of learning complex sets of non-linear features hence rotation invariance is usually achieved by including rotated images during model training. In collecting images of species, a degree of control over the attitude of the subject can be applied however, a degree of unknown rotation is unavoidable. Marker images were rotated between -30° and 30° in increments of 1° degree prior to detection and TL estimation. TL estimates were corrected based on the typical geometry of the detection box in relation to the snout and caudal vertices, as previously described (Figure 8-8). Images were also horizontally flipped. A pyramid of images of decreasing resolution was created by applying a constant scaling factor of 1.5, limited by a minimum resolution of a height or width of 50 pixels. So, let $\boldsymbol{R}$ be a vector representation of image height and width, then the new resolution is $\boldsymbol{R} \cdot {}^{1}/_{1.5}$. RoIs were generated programmatically for scaled, flipped translations and rotated from the original untransformed RoIs (code snippets in geom.py). Inference was then run against these images. Note that rotating a ground truth RoI through θ° does not give the ground truth RoI of an image rotated through θ°, hence IoU is not reported for rotated images.

REMOVING OUTLIERS AND MODELLING BIAS

Poor TL estimates arising from scaling or poor quality images can be easily excluded from final data sets prior to inference according to various metrics (e.g. resolution, blur and lightness), or *a posteriori* by the size of the detection RoI however, this will not control for rotated subjects. Two predictors of the accuracy of TL estimates available after prediction are the width and height of the detection (*size ratio*, an intuitive predictor of rotation) and the objectness score. These two predictors were used to produce a model to reduce TL bias under image rotation without *a priori* knowledge of that rotation. In brief, NASNet CNN detections were split into train and test data. The training data was used to identify biased outliers using an isolation forest (Liu et al., 2008; Pedregosa et al., 2011). Outliers were then removed from the training set, and a gradient boost regressor (Friedman, 2002; Pedregosa et al., 2011) was used to model bias with the predictors size ratio and objectiveness score. Outliers were then removed from the test dataset and the gradient boost model applied to predict bias from objectiveness score and size ratio. Further details are given in Appendix D.

Appendix K. Overview of the European sea bass (*Dicentrarchus labrax*)

BIOLOGY

A comprehensive treatment of the species is given in the reviews of Pickett and Pawson (1994) and Vázquez and Muñoz-Cueto (2014). The European bass is an iteroparous oviparous batch spawner and is gonochronistic (Murua and Saborido-Rey 2003). Sex determination is polygenic and affected by temperature (Piferrer *et al.* 2005, Vandeputte *et al.* 2007) and the sexes display dimorphism in size and growth rates (Saillant *et al.* 2001). Tagging by Pawson and Pickett (1996) of 2205 specimens taken over 8 years provided morphological and maturity data for UK Bass and the following information references his treatment unless otherwise specified. Bass gonadal maturation is strongly associated with length rather than age, total length (TL) at first maturity is 32 cm and 42 cm for males and females respectively, with maturity between 4 and 7 years. Females have significantly greater length at age for ages > 4 years and a separate study by Saillant *et al.* (2001) estimated female weight at ≈ 26% greater than males in fish > 2 years (with some yearly fluctuation). Growth was isometric in both sexes (adjusting for intra-year condition variations), with condition for mature fish north of the Severn estuary at a maximum between September and December which then decreases over the breeding season ending around June.

Bass are widely distributed throughout the coastal waters of Europe, ranging as far north as the Norwegian Sea, down to Southern Morocco and throughout the Mediterranean Sea (Fritsch, Morizur, Lambert, Bonhomme, & Guinand, 2007; Pawson, Pickett, Leballeur, Brown, & Fritsch, 2007b). Sea bass occur over a broad range of habitats as indicated by the varied contents of their stomach (Kelley, 1987). Pawson *et al*. (2007) suggested that climate change has extended the northerly range of sea bass. The tagging study of Kelley (1979) between 1971 to 1975 (n = 912) showed different migratory behaviours between juveniles (< 32 cm) and adults (> 42 cm), with adults migrating to spawning grounds offshore of South Cornwall starting in November and returning through May and June. Juveniles do not migrate to breeding grounds and are thought to stay relatively close to tagging points (Pickett *et al.* 2004).

IMPORTANCE AND MANAGEMENT

Sea bass are particularly vulnerable to over exploitation because of the formation of large spawning aggregations, the relatively long time to reach sexual maturity (females, ~ 6 years) and slow growth rate. However, sea bass had become increasingly important as a commercial species as evidenced by the generally increasing trend in landings between 1985 and 2013 (ICES, 2018) but recruitment has been poor since 2008 (ICES, 2018). Sea bass are prosecuted throughout their range and commercial capture methods include pelagic pair trawling (targeting offshore spawning aggregations), bottom trawling, fixed and drifting gill netting, long lining, seine netting and active line fishing.

Pressure on sea bass stocks are also exerted by recreational fishing, with approximately 25% of all harvest attributed to anglers (Armstrong et al., 2013a; Herfaut et al., 2010; Hyder et al., 2018). Sea bass are highly valued by recreational fishers who frequently report them as their preferred target species in several EU countries. Some countries have also imposed complete bans on commercial fishing of sea bass to maintain the quality of the recreational fishery (e.g. Irish Parliament, 2006; Isle of Man Government, 2016).

In 2014, ICES estimated that spawning biomass had approached the limit reference point (ICES, 2014a) which resulted in the implementation of harvest controls across the EU. Harvest controls were applied to both the commercial and recreational fishery and measures included the increase of the minimum conservation reference size (formerly the minimum landing size) to 42 cm across EU member states, along with additional métier specific temporal closures. Recognising the significant mortality attributed to recreational fishing activity, recreational fishers have had harvest limits applied, with a complete ban on retaining catch between January and September 2018. In 2018 ICES reports that the spawning stock biomass of the North East Atlantic stock in divisions 4.b–c, 7.a, and 7.d–h is below safe biological limits and it appears very likely that harvest controls will remain in place to control mortality induced by the commercial and recreational sectors.

Appendix L. Bias correction modelling

The NASNet (Zoph & Le, 2017) CNN had fewer outliers than the single shot MobileNet (A. G. Howard et al., 2017) and ResNet CNNs (He et al., 2016), hence all the following data is restricted to NASNet detections only. This process describes mean bias error under experimental rotational displacement of the subject along its longitudinal axis from ~0° to ±30° was corrected. The process is described below, and summarised in Figure 8-9.

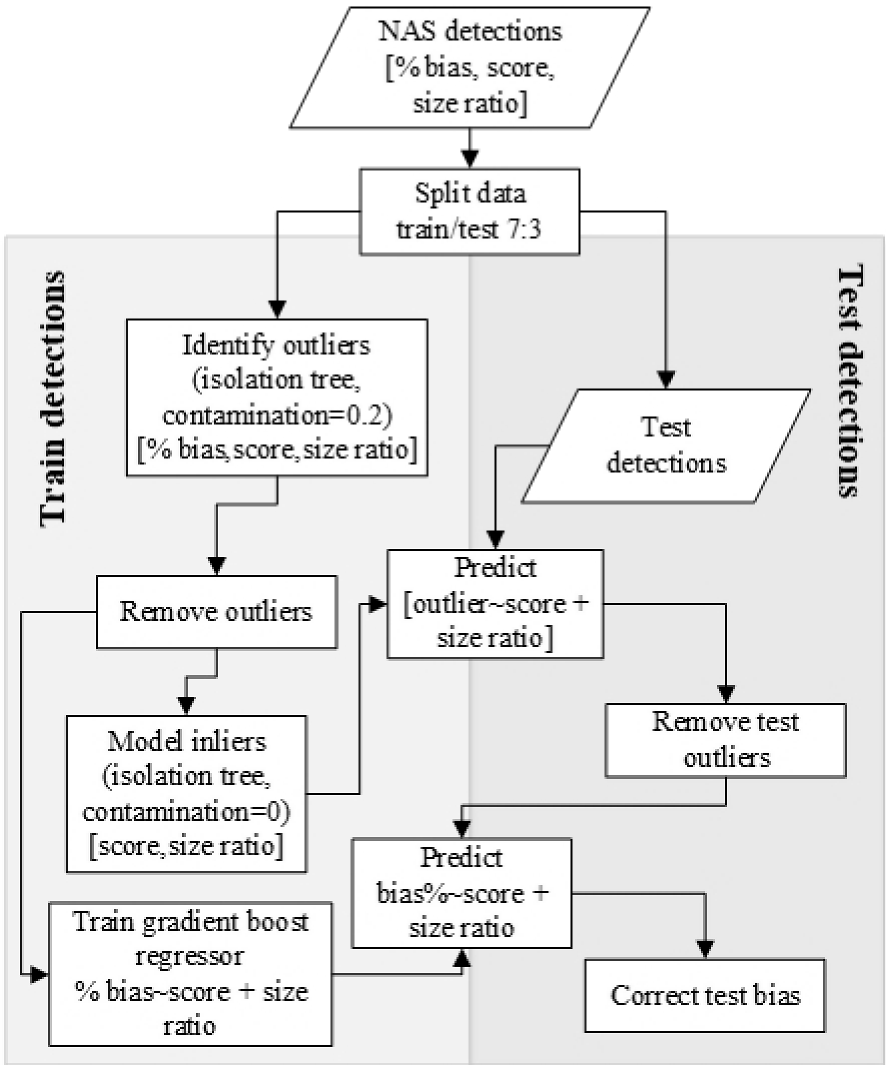

Figure 8-9. Percent bias correction process. Size ration is the ratio of height to width of the detected bounding box. Score is the objectness score calculated by the neural network, which is interpreted as the probability that the proposed region contains the predicted class.

Detections for all rotations were divided randomly into training and test sets (ratio 7:3, train n = 11,555, test n = 4953) with each sample represented by a 3-dimenional vector of percent bias, height-weight ratio and objectiveness score. Percent bias is $\%\ bias = 100 \cdot (\widehat{tl} - tl)/tl$, where tl is the actual total length from manual measurement and $\widehat{tl}$ is the machine vision derived estimate of the total length (rectangle width) after all corrections, as outlined in Appendix B. The height and width ratio is the ratio of sides in pixels of the MV detection rectangle and the objectness score calculated by the neural network, which is interpreted as the probability that the proposed region contains the predicted class.

An isolation forest (Liu et al., 2008; Pedregosa et al., 2011) was trained (contamination = 0.2) to identify outliers (henceforth, *full model*). Contamination means approximately 20% of points are considered as outliers. This was determined by reviewing scatter plots of the data. 20% outliers for the training data for vectors height/width ratio *vs*. % bias, and height/width ratio *vs*. objectiveness score are shown in Figure 8-10 and Figure 8-11 respectively.

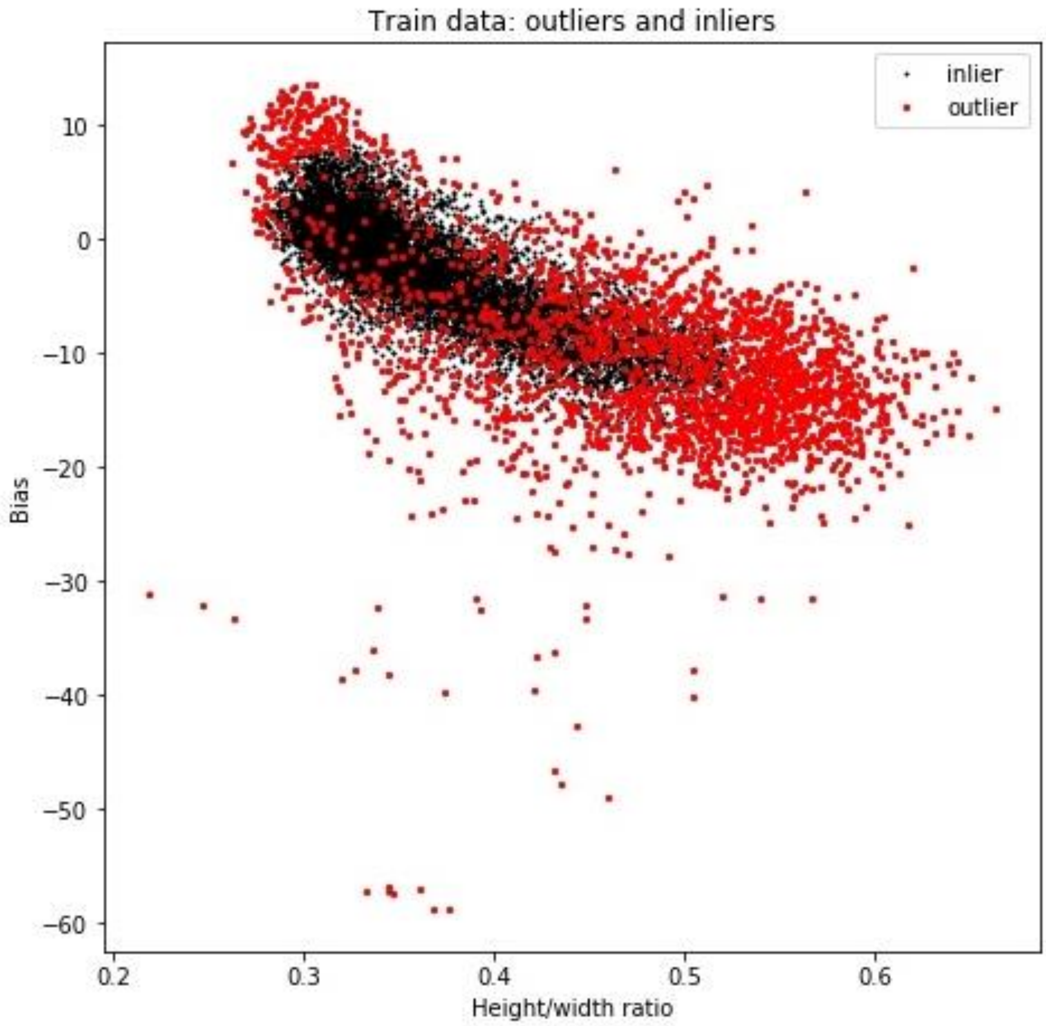

Figure 8-10. Train data outliers (■) and inliers (•) predicted by an isolation forest (Liu et al., 2008; Pedregosa et al., 2011) at 20% contamination for vector axes height/width ratio and % bias.

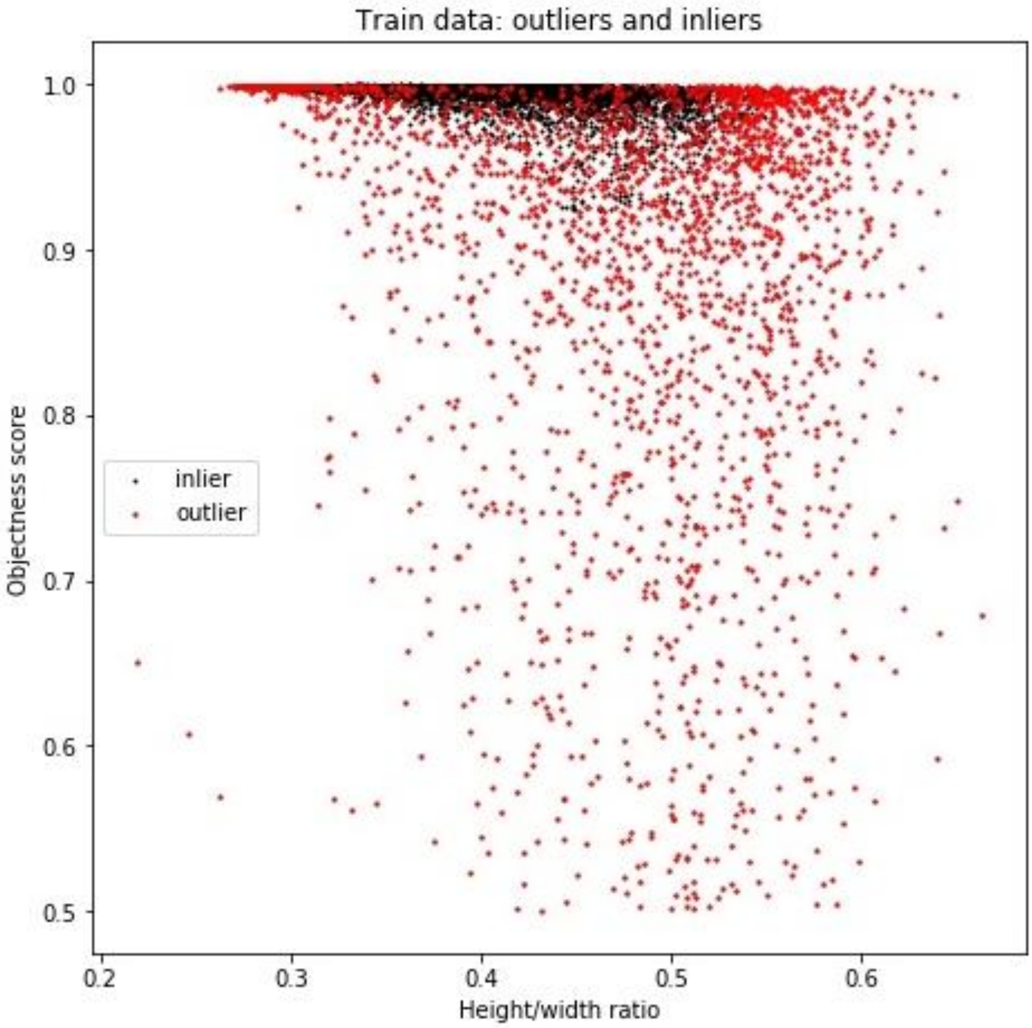

Figure 8-11. Train data outliers (■) and inliers (•) predicted by an isolation forest (Liu et al., 2008; Pedregosa et al., 2011) at 20% contamination for vector axes height/width ratio and objectness score.

Outliers identified by the full model were removed from the training set (*inlier training set*) prior to using a gradient boosting regressor (Friedman, 2002; Pedregosa et al., 2011) with "out of bag" estimator hyper parameter optimization to predict percent bias from height-weight ratio and objectiveness score (*GBR model*).

The bias vector was removed from inlier training set (leaving an *n* by 2 vector of height-weight ratio and objectiveness scores) and this was used to train a second isolation forest (contamination = 0.0) which was subsequently used to remove outliers from the evaluation dataset (i.e. under the assumption of no prior knowledge of total length bias). 293 outliers were identified in the test dataset, with 4660 inliers remaining.

The GBR model was applied to the test data to predict % bias corrections for TL estimates (predictors, height-width ratio and objectiveness score) and these results were reported to determine how effectively bias can be reduced by pruning and correction for results with no other a-priori analysis of the images. The predicted % bias corrections along the axis of the two predictors appear in Figure 8-12 and Figure 8-13.

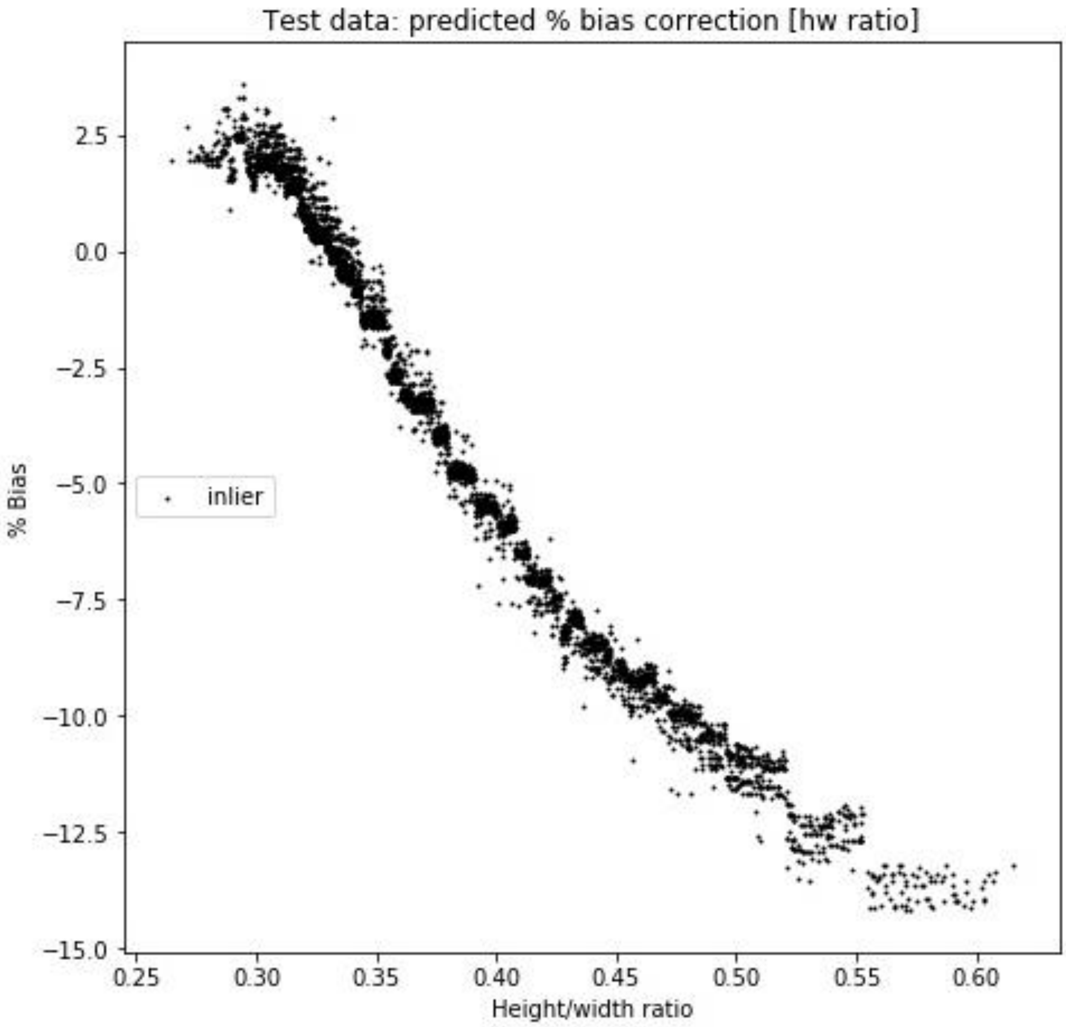

Figure 8-12. Test data predicted % bias correction *vs*. height/width ratio.

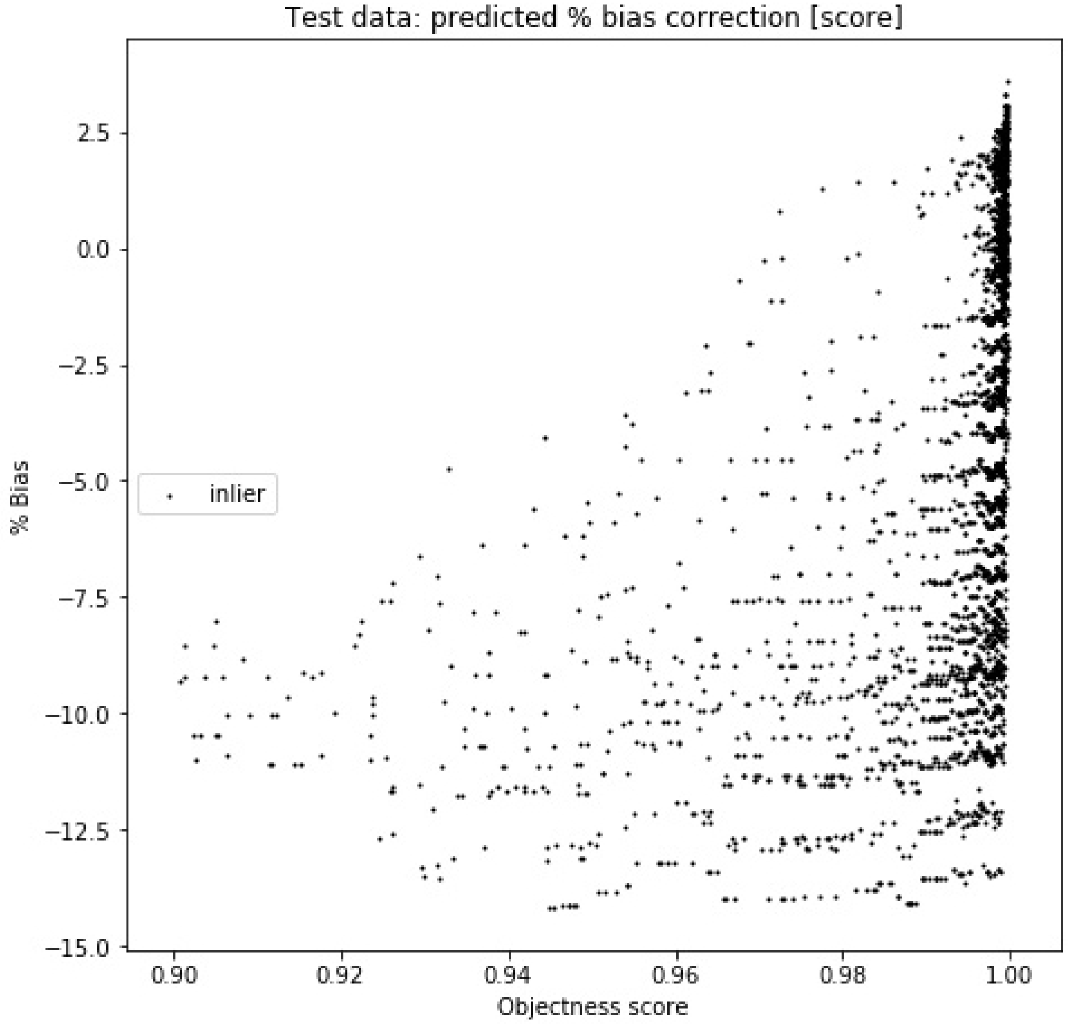

Figure 8-13. Test data predicted % bias correction *vs*. objectness score.

9 Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Web Page]. Retrieved March 10, 2018, from https://www.tensorflow.org/

Addinsoft. (2014). XLSTAT - Essential data analysis tools for Excel [Web Page]. Retrieved June 1, 2016, from https://www.xlstat.com/en

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). Finding high-quality content in social media. *Proceedings of the International Conference on Web Search and Web Data Mining - WSDM '08*, 183. https://doi.org/10.1145/1341531.1341557

Alexander, D. E. (2014). Social media in disaster risk reduction and crisis management. *Sci Eng Ethics*, *20*, 717–733. https://doi.org/10.1007/s11948-013-9502-z

Altieri, A. H., Bertness, M. D., Coverdale, T. C., Herrmann, N. C., Angelini, C., Ecology, S., … Angelini, C. (2012). A trophic cascade triggers collapse of a salt-marsh ecosystem with intensive recreational fishing. *Ecology*, *93*(6), 1402–1410. https://doi.org/https://doi.org/10.1890/11-1314.1

Amazon. (2017a). Amazon Machine Learning [Web Page]. Retrieved February 5, 2017, from https://aws.amazon.com/machine-learning

Amazon. (2017b). Amazon Mechanical Turk: Artificial Intelligence [Web Page]. Retrieved March 2, 2017, from https://www.mturk.com/mturk/welcome

Angling Trust. (2015). Just Take 5 - Help Stop Litter Spoiling Our Beaches And Coastline [Web Page]. Retrieved June 1, 2018, from http://anglingtrust.net/page.asp?section=796&sectionTitle=Help+Stop+Litter+Spoiling+Our+Beaches+And+Coastline

Anuchiracheeva, S., Demaine, H., Shivakoti, G. P., Ruddle, K., Shivakoti, P., & Ruddle, K. (2003). Systematizing local knowledge using GIS: Fisheries management in Bang Saphan Bay, Thailand. *Ocean and Coastal Management*, *46*(11–12), 1049–1068. https://doi.org/10.1016/j.ocecoaman.2004.01.001

Arlinghaus, R. (2006). Understanding recreational angling participation in Germany: Preparing for demographic change. *Human Dimensions of Wildlife*, *11*(4), 229–240. https://doi.org/10.1080/10871200600802889

Arlinghaus, R., & Cowx, I. G. (2002). Reconciling traditional inland fisheries management and sustainability in industrialized countries with emphasis on Europe. *Fish and Fisheries*, *3*, 261–316.

Arlinghaus, R., Tillner, R., & Bork, M. (2015). Explaining participation rates in recreational fishing across industrialised countries. *Fisheries Management and Ecology*, *22*(1), 45–55. https://doi.org/10.1111/fme.12075

Armitage, D. R., Plummer, R., Berkes, F., Arthur, R. I., Charles, A. T., Davidson-Hunt, I. J., … Wollenberg, E. K. (2009). Adaptive co-management for social-ecological complexity. *Frontiers in Ecology and the Environment*, *7*(2), 95–102. https://doi.org/10.1890/070089

Armstrong, M., Brown, A., Hargreaves, J., Hyder, K., Munday, M., Proctor, S., & Roberts, A. (2013a). *Sea Angling 2012 – a survey of recreational sea angling activity and economic value in England. Annex 4: An on-site survey of recreational sea angling catches from the shore and from private and rental boats in England in 2012* [Report]. Retrieved from http://webarchive.nationalarchives.gov.uk/20140305101647/http://www.marinemanagement.org.uk/seaangling/finalreport.htm

Armstrong, M., Brown, A., Hargreaves, J., Hyder, K., Munday, M., Proctor, S., & Roberts, A. (2013b). *Sea Angling 2012 – a survey of recreational sea angling activity and economic value in England. Annex 6: An angling tourism study in the Isles of Scilly* [Report]. Retrieved from http://www.marinemanagement.org.uk/seaangling/finalreport.htm

Aron, H., Eno, C., & Jones, B. (2014). *Natural Resources Wales: FishMap Môn Project Report 2014* [Report] (Vol. 2014). NRW. Retrieved from https://1drv.ms/b/s!Au4azAJUYz5JiJBRx37VuZxO6gDJKg

Ashford, J. R., Jones, C. M., & Fegley, L. (2009). A license registry improves sampling efficiency for a marine recreational survey. *Transactions of the American Fisheries Society*, (138), 984–989. https://doi.org/10.1577/T08-142.1

Asoh, K., Yoshikawa, T., Kosaki, R., & Marschall, E. A. (2004). Damage to cauliflower coral by monofilament fishing lines in Hawaii. *Conservation Biology*, *18*(6), 1645–1650. https://doi.org/10.2307/3589046

Association of Internet Researchers. (2017). Association of Internet Researchers (AoIR) - Ethics [Web Page]. Retrieved February 20, 2017, from https://aoir.org/ethics/

Australian Parliament. Copyright Act 1968. Act No. 63 of 1968, ss 40(1), 103C(1) (1968).

Australian Parliament. Privacy Act 1988. Act No. 119 of 1988. (1988). Retrieved from https://www.legislation.gov.au/Series/C2004A03712

Babbie, E. R. (2013). *The practice of social research* (13th ed.). [Book], Belmont CA: Wadsworth Cengage Learning. Retrieved from http://www.worldcat.org/title/practice-of-social-research/oclc/774699583

Banha, F., Veríssimo, A., Ribeiro, F., & Anastácio, P. M. (2017). Forensic reconstruction of *Ictalurus punctatus* invasion routes using on-line fishermen records. *Knowledge & Management of Aquatic Ecosystems*, *418*(418), 56. https://doi.org/10.1051/kmae/2017045

Barbedo, J. G. A. (2017). A new automatic method for disease symptom segmentation in digital photographs of plant leaves. *European Journal of Plant Pathology*, *147*(2), 349–364. https://doi.org/10.1007/s10658-016-1007-6

Barbini, S. A., Lucifora, L. O., Figueroa, D. E., & Gillanders, B. (2015). Using opportunistic records from a recreational fishing magazine to assess population trends of sharks. *Canadian Journal of Fisheries and Aquatic Sciences*, *1859*, 1–7. https://doi.org/10.1139/cjfas-2015-0087

Barrow, W., Brickley, K. W., Dumbrell, E., Johnson, N., & Fisheries and Oceans Canada. (2012). *Survey of recreational fishing in Canada, 2010* [Report]. Ottawa, ON (Canada): Fisheries and Oceans Canada (DFO). https://doi.org/978-1-100-54196-9

Bartholomew, A., & Bohnsack, J. A. (2005). A review of catch-and-release angling mortality with implications for no-take reserves. *Reviews in Fish Biology and Fisheries*, *15*(1–2), 129–154.

Bartholomew, D. C., Mangel, C., Alfaro-shigueto, J., Pingo, S., Jimenez, A., Godley, B. J., … Godley, B. J. (2018). Remote electronic monitoring as a potential alternative to on-board observers in small-scale fisheries. *Biological Conservation*, *219*(May 2017), 35–45. https://doi.org/10.1016/j.biocon.2018.01.003

Barve, V. (2014). Discovering and developing primary biodiversity data from social networking sites: A novel approach. *Ecological Informatics*, *24*, 194–199. https://doi.org/10.1016/j.ecoinf.2014.08.008

Beauchamp, T. L., & Childress, J. F. (2001). *Principles of biomedical ethics* (1st ed.). [Book], Oxford University Press, USA.

Belhabib, D., Campredon, P., Lazar, N., Sumaila, U. R., Baye, B. C., Kane, E. A., & Pauly, D. (2016). Best for pleasure, not for business: evaluating recreational marine fisheries in West Africa using unconventional sources of data. *Palgrave Communications*, *2*(15050), 1–10. https://doi.org/10.1057/palcomms.2015.50

Bellquist, L., & Semmens, B. X. (2016). Temporal and spatial dynamics of ’trophy’-sized demersal fishes off the California (USA) coast, 1966 to 2013. *Marine Ecology Progress Series*, *547*(April), 1–18. https://doi.org/10.3354/meps11667

Bennett, B. A., Attwood, C. G., & Mantel, J. D. (1994). Teleost catches by three shore-angling clubs in the South-western Cape, with an assessment of the effect of restrictions applied in 1985. *South African Journal of Marine Science*, *14*(1), 11–18. https://doi.org/10.2989/025776194784286978

Benoît, H. P., & Allard, J. (2009). Can the data from at-sea observer surveys be used to make general inferences about catch composition and discards? *Canadian Journal of Fisheries and Aquatic Sciences*, *66*(12), 2025–2039. https://doi.org/10.1139/F09-116

Bergamasco, F., Albarelli, A., Rodol, E., Torsello, A., Ambientali, S., Statistica, I., & Venezia, F. (2011). RUNE-Tag : A high accuracy fiducial marker with strong occlusion resilience. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE. https://doi.org/10.1109/CVPR.2011.5995544

Bergeron, P. (2007). Parallel lasers for remote measurements of morphological traits. *The Journal of Wildlife Management*, *71*(1), 289–292. https://doi.org/10.2307/4495176

Beyer, H. L. (2015). Geospatial Modelling Environment | SpatialEcology.Com. [Computer Program], Hawthorne L. Beyer. Retrieved from http://www.spatialecology.com/gme/index.htm

Bian, J., Liu, Y., Zhou, D., Agichtein, E., & Zha, H. (2009). Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *Proceedings of the 18th international conference on World wide web - WWW '09* (p. 51). New York, New York, USA: ACM Press. https://doi.org/10.1145/1526709.1526717

Bicknell, A. W. J., Godley, B. J., Sheehan, E. V., Votier, S. C., & Witt, M. J. (2016). Camera technology for monitoring marine biodiversity and human impact. *Frontiers in Ecology and the Environment*, *14*(8), 424–432. https://doi.org/10.1002/fee.1322

Billfish Foundation. (2018). The Billfish Tag and Release Program [Web Page]. Retrieved May 1, 2018, from https://billfish.org/research/tag-and-release/

Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python: NTLK* (1st ed.). [Book], O'Reilly Media Inc.

Blight, S. J., & Smallwood, C. B. (2015). *Technical manual for camera surveys of boat and shore based recreational fishing in Western Australia. Fisheries Occasional Publication No. 121* [Report]. Retrieved from http://www.fish.wa.gov.au/Documents/occasional_publications/fop121.pdf

Boesea, B. L., Clinton, P. J., Dennis, D., Golden, R. C., & Kim, B. (2008). Digital image analysis of *Zostera marina* leaf injury. *Aquatic Botany*, *88*(1), 87–90. https://doi.org/10.1016/J.AQUABOT.2007.08.016

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, *489*(7415), 295–298. https://doi.org/10.1038/nature11421

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next steps for citizen science. *Science*, *343*(6178), 1436–1437. https://doi.org/10.1126/science.1251554

Bouchet, P. J., & Meeuwig, J. J. (2015). Drifting baited stereo-videography: a novel sampling tool for surveying pelagic wildlife in offshore marine reserves. *Ecosphere*, *6*(8), art137. https://doi.org/10.1890/ES14-00380.1

Boyes, S. J., & Elliott, M. (2014). Marine legislation - The ultimate "horrendogram": International law, European directives & national implementation. *Marine Pollution Bulletin*, *86*(1–2), 39–47. https://doi.org/10.1016/j.marpolbul.2014.06.055

Boyes, S. J., Elliott, M., Thomson, S. M., Atkins, S., & Gilliland, P. (2007). A proposed multiple-use zoning scheme for the Irish Sea. An interpretation of current legislation through the use of GIS-based zoning approaches and effectiveness for the protection of nature conservation interests. *Marine Policy, 31*(3), 287–298. https://doi.org/10.1016/j.marpol.2006.08.005

Branson, S., Van Horn, G., Belongie, S., & Perona, P. (2014). Improved bird species categorization using pose normalized deep convolutional nets. In *British Machine Vision Conference*. Retrieved from http://arxiv.org/abs/1406.2952

Brennan, J., Fitzsimmons, C., Gray, T., & Raggatt, L. (2014). EU marine strategy framework directive (MSFD) and marine spatial planning (MSP): Which is the more dominant and practicable contributor to maritime policy in the UK? *Marine Policy*, *43*, 359–366. https://doi.org/10.1016/j.marpol.2013.07.011

Brick, J. M., Andrews, W. R., & Mathiowetz, N. (2012). *A Comparison of Recreational Fishing Effort Survey Designs* [Report]. Retrieved from https://www.st.nmfs.noaa.gov/mdms/doc/08A_Comparison_of_Fishing_Effort_Surveys_Report_FINAL.pdf

Brown, A. (2012). *The National Angling Survey 2012: Survey Report* [Report]. Manchester: Substance. Retrieved from http://www.resources.anglingresearch.org.uk/sites/resources.anglingresearch.org.uk/files/National_Angling_Survey_Report_2012.pdf

Brown, A., Djohari, N., & Stolk, P. (2012). *Fishing for Answers: The Final Report of the Social and Community Benefits of Angling Project* [Report]. Manchester, UK: Substance. Retrieved from http://resources.anglingresearch.org.uk/project_reports/final_report_2012

Brown, A., Munday, M., Roberts, A., Roche, N., Armstrong, M., Hyder, K., & Oughton, E. (2013). *Sea Angling 2012 – A survey of recreational sea angling activity and economic value in England. Annex 2: The Economic and Social Value of Recreational Sea Angling in England* [Report]. Marine Management Organisation. Retrieved from http://www.marinemanagement.org.uk/seaangling/finalreport.htm

Brown, A., Stolk, P., & Dojhari, N. (2010). *Angling: A social research overview* [Report]. Manchester: Substance. Retrieved from http://resources.anglingresearch.org.uk/sites/resources.anglingresearch.org.uk/files/Substance-EA-Angling-Research-Review.pdf

Brown, K. F., Long, S. J., Ramsay, M., Hudson, M. J., Green, J., Vincent, C. A., … Sevdalis, N. (2012). UK parents' decision-making about measles–mumps–rubella (MMR) vaccine 10 years after the MMR-autism controversy: A qualitative analysis. *Vaccine*, *30*(10), 1855–1864.

https://doi.org/10.1016/j.vaccine.2011.12.127

Brunelli, R. (2009). *Template matching techniques in computer vision : Theory and practice*. [Book], Wiley.

Cadrin, S. X., & Friedland, K. D. (1999). The utility of image processing techniques for morphometric analysis and stock identication. *Fish. Res.*, *43*, 129–139. https://doi.org/10.1016/S0165-7836(99)00070-3

Campbell, D., West, L., Lyle, J., McGlennon, D., Coleman, A., Henry, G., & Reid, D. (2001). *The Australian National Recreational Fishing Survey: 2000-01* [Report]. Australia: IIFET.

Canadian Parliament. Copyright Act, R.S.C., 1985, c. C-42 s. 29. Exeptions - Fair Dealing (1985).

Canadian Parliament. Privacy Act (R.S.C., 1985, c. P-21) (1985).

Canese, S., & Bava, S. (2015). The decline of top predators in deep coral reefs. In *Proceedings of the 1st Mediterranean Symposium on the Conservation of Dark Habitats* (pp. 67–68). Portorož, Slovenia: UNEP. Retrieved from http://rac-spa.org/sites/default/files/symposium/proc_1_dark_habitats_final.pdf#page=76

Carpenter, S. R., Munozdelrio, A., Newman, S., Rasmussen, P. W., & Johnson, B. M. (1994). Interactions of anglers and walleyes in Escanaba Lake, Wisconsin. *Ecological Applications*, *4*(4), 822–832. https://doi.org/10.2307/1942011

CEFAS. (2012). Sea Angling 2012 - Shore and private boat survey. Data from shore and private boat survey conducted by IFCAs within the Sea Angling 2012 project [Web Page]. Retrieved November 19, 2017, from http://www.marinemanagement.org.uk/seaangling/finalreport.htm

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science (New York, N.Y.)*, *329*(5996), 1194–1197. https://doi.org/10.1126/science.1185231

Chambers, C., Haines, R., & Pitts, J. (2013). *Wales Activity Mapping: Economic Valuation of Marine Recreation Activity Final Report* [Report]. London: Marine Planning Consultants. Retrieved from http://www.walesactivitymapping.org.uk/wp-content/uploads/2013/07/Wales-Activity-Mapping-Economic-Valuation-of-Marine-Recreation-Activity-Full-Report-Nov-2013.pdf

Chang, S.-K., DiNardo, G., & Lin, T.-T. (2010). Photo-based approach as an alternative method for collection of albacore (*Thunnus alalunga*) length frequency from longline vessels. *Fisheries Research*, *105*(3), 148–155. https://doi.org/10.1016/J.FISHRES.2010.03.021

Chang, S.-K., Lin, T. Te, Lin, G. H., Chang, H. Y., & Hsieh, C. L. (2009). How to collect verifiable length data on tuna from photographs: An approach for sample vessels. *ICES Journal of Marine Science*, *66*(5), 907–915. https://doi.org/10.1093/icesjms/fsp108

Chiappone, M., Dienes, H., Swanson, D. W., & Miller, S. L. (2005). Impacts of lost fishing gear on coral reef sessile invertebrates in the Florida Keys National Marine Sanctuary. *Biological Conservation*, *121*(2), 221–230. https://doi.org/10.1016/j.biocon.2004.04.023

Chuang, M. C., Hwang, J. N., Williams, K., & Towler, R. (2011). Automatic fish segmentation via double local thresholding for trawl-based underwater camera systems. *Proceedings - International Conference on Image Processing, ICIP*, 3145–3148. https://doi.org/10.1109/ICIP.2011.6116334

Claassens, L., & Hodgson, A. N. (2018). Gaining insights into in situ behaviour of an endangered seahorse using action cameras. *Journal of Zoology*, *304*(2), 98–108. https://doi.org/10.1111/jzo.12509

Claireaux, G., & Lagardère, J.-P. (1999). Influence of temperature, oxygen and salinity on the metabolism of the European sea bass. *Journal of Sea Research*, *42*(2), 157–168. https://doi.org/10.1016/S1385-1101(99)00019-2

Close, C. H., & Hall, G. B. (2006). A GIS-based protocol for the collection and use of local knowledge in fisheries management planning. *Journal of Environmental Management*, *78*(4), 341–352. https://doi.org/10.1016/j.jenvman.2005.04.027

COCO. (2018). COCO - Common Objects in Context [Web Page]. Retrieved August 1, 2018, from http://cocodataset.org/#home

Cohen, J. (1968). Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19673146

Coleman, F. C., Figueira, W. F., Ueland, J. S., & Crowder, L. B. (2004). The impact of United States recreational fisheries on marine fish populations. *Science*, *305*(5692), 1958–1960. https://doi.org/10.1126/science.1100397

Cooke, S. J., & Cowx, I. G. (2004). The role of recreational fishing in global fish crises. *Bioscience*, *54*(9), 857–859. https://doi.org/10.1641/0006-3568(2004)054[0857:TRORFI]2.0.CO;2

Costa, C., Loy, A., Cataudella, S., Davis, D., & Scardi, M. (2006). Extracting fish size using dual underwater cameras. *Aquacultural Engineering*, *35*(3), 218–227. https://doi.org/10.1016/J.AQUAENG.2006.02.003

Costello, C., Ovando, D., Hilborn, R., Gaines, S. D., Deschenes, O., & Lester, S. E. (2012). Status and solutions for the world's unassessed fisheries. *Science*, *338*, 517–520. https://doi.org/10.1126/science.1223389

Cox, K. (2015). *Issue Brief: Text and Data Mining and Fair Use in the United States* [Report]. Retrieved from http://www.arl.org/storage/documents/TDM-5JUNE2015.pdf

Crozier, G. K. D., & Schulte-Hostedde, A. I. (2015). Towards improving the ethics of ecological research. *Science and Engineering Ethics*, *21*(3), 577–594. https://doi.org/10.1007/s11948-014-9558-4

Curtin, S. (2009). Wildlife tourism: the intangible, psychological benefits of human–wildlife encounters. *Current Issues in Tourism*, *12*(5–6), 451–474. https://doi.org/10.1080/13683500903042857

Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* (Vol. I, pp. 886–893). https://doi.org/10.1109/CVPR.2005.177

Dann, S. L., Alvarado, A., Palmer, D., Schroeder, B., & Stephens, M. (2008). *Angler participation, recruitment, and retention in Michigan, 1995-2004: Using data-mining techniques for customer relationship management. Research Report 2088.* [Report]. Retrieved from http://www.dnr.state.mi.us/publications/pdfs/IFR/ifrlibra/Research/reports/2088/2088.pdf

Daume, S., Albert, M., Gadow, K. Von, & von Gadow, K. (2014). Forest monitoring and social media – Complementary data sources for ecosystem surveillance? *Forest Ecology and Management*, *316*, 9–20. https://doi.org/10.1016/j.foreco.2013.09.004

De Longueville, B., Smith, R. S., & Luraschi, G. (2009). OMG, from here, I can see the flames! In *Proceedings of the 2009 International Workshop on Location Based Social Networks - LBSN '09* (p. 73). New York, New York, USA: ACM Press. https://doi.org/10.1145/1629890.1629907

Deakos, M. H. (2010). Paired-laser photogrammetry as a simple and accurate system for measuring the body size of free-ranging manta rays *Manta alfredi*. *Aquatic Biology*, *10*(1), 1–10. https://doi.org/10.3354/ab00258

Deepananda, K. H. M. A., Amarasinghe, U. S., Jayasinghe-Mudalige, U. K., & Berkes, F. (2016). Stilt fisher knowledge in southern Sri Lanka as an expert system: A strategy towards co-management. *Fisheries Research*, *174*, 288–297. https://doi.org/http://dx.doi.org/10.1016/j.fishres.2015.10.028

Di Camillo, C. G., Ponti, M., Bavestrello, G., Krzelj, M., & Cerrano, C. (2018). Building a baseline for habitat-forming corals by a multi-source approach, including Web Ecological Knowledge. *Biodiversity and Conservation*, *27*(5), 1257–1276. https://doi.org/10.1007/s10531-017-1492-8

Di Minin, E., Fink, C., Hiippala, T., & Tenkanen, H. (2018). Use of machine learning to investigate illegal wildlife trade on social media. *Conservation Biology*. https://doi.org/10.1111/cobi.13104

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*, 63.

https://doi.org/10.3389/fenvs.2015.00063

DigiKam. (2018). digiKam, Open Source Professional Photo Management. [Computer Program], The digiKam team. Retrieved from https://www.digikam.org/about/

Ding, J., Chen, B., Liu, H., & Huang, M. (2016). Convolutional neural network with data augmentation for SAR target recognition. *IEEE Geoscience and Remote Sensing Letters*, *13*(3), 364–368.

Donnelley, R. R., Radford, A., Riddington, R., & Gibson, H. G. (2009). *Technical Report: Economic Impact of Recreational Sea Angling in Scotland* [Report] (Vol. SAG/0005/0). Edinburgh: Scottish Government. Retrieved from https://www.gov.scot/Publications/2009/07/31154700/0

Douvere, F. (2008). The importance of marine spatial planning in advancing ecosystem-based sea use management. *Marine Policy*, *32*(5), 762–771. https://doi.org/10.1016/J.MARPOL.2008.03.021

Drew Associates Ltd. (2004). *Research into the Economic Contribution of Sea Angling commissioned by the Economics and Statistics Group of the Department for Environment, Food and Rural Affairs (DEFRA)* [Report]. Drew Associates. Retrieved from Not Available

Dugas, A. F., Jalalpour, M., Gel, Y., Levin, S., Torcaso, F., Igusa, T., & Rothman, R. E. (2013). Influenza forecasting with Google Flu Trends. *PLoS ONE*, *8*(2), e56176. https://doi.org/10.1371/journal.pone.0056176

Dulvy, N. K., Freckleton, R. P., & Polunin, N. V. C. (2004). Coral reef cascades and the indirect effects of predator removal by exploitation. *Ecology Letters*, *7*(5), 410–416. https://doi.org/10.1111/j.1461-0248.2004.00593.x

Dunbrack, R. L. (2006). In situ measurement of fish body length using perspective-based remote stereo-video. *Fisheries Research*, *82*(1–3), 327–331. https://doi.org/10.1016/J.FISHRES.2006.08.017

Dutta, A., Gupta, A., & Zisserman, A. (2018). VGG Image Annotator. [Computer Program], University of Oxford, Visual Geometry Group. Retrieved from http://www.robots.ox.ac.uk/~vgg/software/via/

E T Spatial Technologies. (2014). ET GeoWizards 11.2. [Computer Program], ET Spatial Technologies. Retrieved from http://www.ian-ko.com/

Eero, M., Strehlow, H. V., Adams, C. M., & Vinther, M. (2014). Does recreational catch impact the TAC for commercial fisheries? *ICES Journal of Marine Science*, *72*(2), 450–457. https://doi.org/10.1093/icesjms/fsu121

Eid, E., & Handal, R. (2017). Illegal hunting in Jordan: using social media to assess impacts on wildlife. *Oryx*, 1–6. https://doi.org/10.1017/S0030605316001629

EIFAC. (2008). *European Inland Fisheries Advisory Commission (EIFAC) code of practice for recreational fisheries. EIFAC Occasional Paper 42.* [Report]. Rome. Retrieved from http://www.fao.org/docrep/012/i0363e/i0363e00.htm

Environ, F. E., Galaz, V., Crona, B., Daw, T., Bodin, Ö., Nyström, M., & Olsson, P. (2010). Can web crawlers revolutionize ecological monitoring? *Frontiers in Ecology and the Environment*, *8*(2), 99–104. https://doi.org/10.1890/070204

Environment Agency. (2014, September). New online service to help monitor salmon and sea trout stocks [Web Page]. Retrieved April 1, 2016, from https://environmentagency.blog.gov.uk/2014/09/01/new-online-service-to-help-monitor-salmon-and-sea-trout-stocks/

Environmental Law Institute. (2009). *Ocean and Coastal Ecosystem-Based Management: Implementation Handbook* [Report]. Environmental Law Institute. Retrieved from https://www.eli.org/sites/default/files/eli-pubs/d19_03.pdf

ESRI. (2010). ArcMap 10.0. [Computer Program], ESRI. Retrieved from http://desktop.arcgis.com/en/arcmap

European Commission. Directive 92/43/EEC: Birds and Habitats Directive (1992). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31992L0043

European Commission. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 On the protection of individuals with regard to the processing of personal data and on the free movement of such data (1995).

European Commission. Charter of Fundamental Rights of the European Union (2000/C 364/01): Article 8, Protection of Personal Data (2000).

European Commission. Directive 2001/29/EC of the European Parliament and of The Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society, Pub. L. No. 5.3(a), Official Journal of the European Union 1 (2001). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32001L0029&from=EN

European Commission. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008: Establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive) (2008). Official Journal of the European Union.

European Commission. Council Regulation (EC) No. 1224/2009 of November 2009. Establishing a Community control system for ensuring compliance with the rules of the Common Fisheries Policy. L343 (2009).

European Commission. Communication from the Commission to the Europeam Parliament, the Council, the Economic and Social Committee and the Committee of the Regions: ‘A comprehensive approach on personal data protection in the European Union’ (2010).

European Commission. Commission Decision: Adopting a multiannual Community programme for the collection , management and use of data in the fisheries sector for the period 2011-2013, 2010/93/EU, Official Journal of the European Union § (2013). European Union. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32010D0093

European Commission. Directive 2014/89/EU of the European Parliament and of the Council of 23 July 2014 establishing a framework for maritime spatial planning (2014). EC.

European Commission. Commission Implementing Regulation (EU) 2015/1316 of 30 July 2015 derogating from Council Regulation (EC) No 850/98, as regards the minimum conservation reference size for sea bass (Dicentrarchus labrax) (2015). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1440763073745&uri=CELEX:32015R1316

European Commission. Council Regulation (EU) 2015/523 of 25 March 2015 amending Regulations (EU) No 43/2014 and (EU) 2015/104 as regards certain fishing opportunities, Official Journal of the European Union § (2015). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32015R0523

European Commission. Commission Implementing Decision (EU) 2016/1251 of 12 July 2016 adopting a multiannual Union programme for the collection, management and use of data in the fisheries and aquaculture sectors for the period 2017-2019 (notified under document C(2016) 4329), Official Journal of the European Parliament § (2016). European Union. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.207.01.0113.01.ENG

European Commission. Council Regulation (EU) 2016/72 of 22 January 2016 fixing for 2016 the fishing opportunities for certain fish stocks and groups of fish stocks, applicable in Union waters and, for Union fishing vessels, in certain non-Union waters, and amending Regulation (2016). Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/865aa5e4-c593-11e5-a4b5-01aa75ed71a1/language-en

European Commission. Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market. COM(2016) 593 final. (2016).

European Commission. Regulation (EU) 2016/679 of the European Parliament of the Council of 27 April 2016 On the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (2016). Official Journal of the European Union. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2016.119.01.0001.01.ENG&toc=OJ:L:2016:119:TOC

European Commission. COMMISION DECISION (EU) 2017/848 of 17 May 2017 laying down criteria and methodological standards on good environmental status of marine waters and specifications and standarised methods for monitoring and assessment, and repealing Decision 2010/477/EU (2017). Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017D0848&from=EN

European Commission. Regulation (EU) 2017/1004 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 17 May 2017 on the establishment of a Union framework for the collection, management and use of data in the fisheries sector and support for scientific advice regarding the common (2017). Retrieved from https://datacollection.jrc.ec.europa.eu/c/document_library/get_file?uuid=170c3f7d-f54d-470d-85b7-c3f65edc0097&groupId=10213

European Commission. COUNCIL REGULATION (EU) 2018/120 of 23 January 2018 fixing for 2018 the fishing opportunities for certain fish stocks and groups of fish stocks, applicable in Union waters and, for Union fishing vessels, in certain non-Union waters, and amending Regulatio (2018). Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018R0120&from=EN

Eysenbach, G. (2011). Infodemiology and Infoveillance. *American Journal of Preventive Medicine*, *40*(5), S154–S158. https://doi.org/10.1016/j.amepre.2011.02.006

Fairclough, D. V, Brown, J. I., Carlish, B. J., Crisafulli, B. M., & Keay, I. S. (2014). Breathing life into fisheries stock assessments with citizen science. *Scientific Reports*, *4*, 7249. https://doi.org/10.1038/srep07249

Faunce, C. H., & Barbeaux, S. J. (2011). The frequency and quantity of Alaskan groundfish catcher-vessel landings made with and without an observer. *ICES Journal of Marine Science*, *68*(8), 1757–1763. https://doi.org/10.1093/icesjms/fsr090

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., & Ramanan, D. (2010). Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(9), 1627–1645. https://doi.org/10.1109/MC.2014.42

Ferter, K., Weltersbach, M. S., Strehlow, H. V., Vølstad, J. H., Alós, J., Arlinghaus, R., … Dupont, S. (2013). Unexpectedly high catch-and-release rates in European marine recreational fisheries: implications for science and management. *ICES Journal of Marine Science*, *70*(7), 1319–1329. https://doi.org/10.1093/icesjms/fst104

Fishbrain. (2018). Fishbrain [Web Page]. Retrieved July 19, 2018, from https://fishbrain.com/mission/

Flannery, W., Healy, N., & Luna, M. (2018). Exclusion and non-participation in marine spatial planning. *Marine Policy*, *88*(January), 32–40. https://doi.org/10.1016/j.marpol.2017.11.001

Florisson, J. H., Tweedley, J. R., Walker, T. H. E., & Chaplin, J. A. (2018). Reef vision: A citizen science program for monitoring the fish faunas of artificial reefs. *Fisheries Research*, *206*(June), 296–308. https://doi.org/10.1016/j.fishres.2018.05.006

Forbes. (2018). How Much Data Do We Create every day? The Mind-Blowing Stats Everyone Should Read [Web Page]. Retrieved August 10, 2018, from https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#184a642460ba

Fowler Jr, F. J., & Mangione, T. W. (1990). *Standardized survey interviewing: Minimizing interviewer-related error* (Vol. 18). [Book], Sage.

Fox, J., & Weisberg, S. (2011). *An {R} Companion to Applied Regression* (Second). [Book], Thousand Oaks {CA}: Sage. Retrieved from http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Frankel, M. S., & Siang, S. (1999). *Ethical and legal aspects of human subjects research on the Internet* [Report]. Washington D.C. Retrieved from https://nationalethicscenter.org/resources/187/download/ethical_legal.pdf

Freese, C. (2012). *Wild species as commodities: managing markets and ecosystems for sustainability*. [Book], Island Press.

Freitas, M. De, Roberto, P., Tagliani, A., De Freitas, D. M., & Tagliani, P. R. A. (2009). The use of GIS for the integration of traditional and scientific knowledge in supporting artisanal fisheries management in southern Brazil. *Journal of Environmental Management*, *90*(6), 2071–2080. https://doi.org/10.1016/j.jenvman.2007.08.026

Freudenberg, P., & Arlinghaus, R. (2010). Benefits and constraints of outdoor recreation for people with physical disabilities: Inferences from recreational fishing. *Leisure Sciences*, *32*(1), 55–71. https://doi.org/http://0-dx.doi.org.unicat.bangor.ac.uk/10.1080/01490400903430889

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, *38*(4), 367–378. https://doi.org/10.1016/S0167-9473(01)00065-2

Frijlink, S., & Lyle, J. M. (2010). *An evaluation of motivations, attitudes and awareness of Tasmanian recreational fishers* [Report]. Hobart, Tasmania: Institute for Marine and Antarctic Studies. Retrieved from https://dpipwe.tas.gov.au/Documents/Motivation-Attitude_Report.pdf

Fritsch, M., Morizur, Y., Lambert, E., Bonhomme, F., & Guinand, B. (2007). Assessment of sea bass (*Dicentrarchus labrax*, L.) stock delimitation in the Bay of Biscay and the English Channel based on mark-recapture and genetic data. *Fisheries Research*, *83*(2–3), 123–132. https://doi.org/10.1016/j.fishres.2006.09.002

Garcia, S. M. (1994). The Precautionary Principle: Its implications in capture fisheries management. *Ocean & Coastal Management*, *22*, 99–125. https://doi.org/10.1016/0964-5691(94)90014-0

Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F. J., & Marín-Jiménez, M. J. (2014). Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, *47*(6), 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005

Gartner, W. C., Love, L. L., & Erkkila, D. (2002). *Economic Impact and Social Benefits Study of Coldwater Angling in Minnesota* [Report]. Minnesota: Tourism Center, University of Minnesota Extension Service. Retrieved from https://conservancy.umn.edu/handle/11299/170195

Gartside, D. F., Harrison, B., & Ryan, B. L. (1999). An evaluation of the use of fishing club records in the management of marine recreational fisheries. *Fisheries Research*, *41*(1), 47–61. https://doi.org/10.1016/S0165-7836(99)00007-7

Georgeson, L., Moore, A., Ward, P., Stenekes, N., Kancans, R., Manzur, K., … Stobutzki, I. (2015). *A framework for regular national recreational fishing surveys* [Report]. Retrieved from https://data.gov.au/dataset/ds-dga-af047dd1-1826-4d1c-9ec7-bda3de2e5021/details

Gill, D. J. C., Fa, J. E., Rowcliffe, J. M., & Kümpel, N. F. (2012). Drivers of change in hunter offtake and hunting strategies in Sendje, Equatorial Guinea. *Conservation Biology*, *26*(6), 1052–1060. https://doi.org/10.1111/j.1523-1739.2012.01876.x

Giovos, I., Keramidas, I., Antoniou, C., Deidun, A., Font, T., Kleitou, P., … Moutopoulos, D. K. (2018). Identifying recreational fisheries in the Mediterranean Sea through social media. *Fisheries Management and Ecology*, *25*(4), 287–295. https://doi.org/10.1111/fme.12293

Giri, K., & Hall, K. (2015). *South Australian Recreational Fishing Survey 2013/14. Internal Report Seres No. 62* [Report]. Melbourne. Retrieved from http://www.pir.sa.gov.au/__data/assets/pdf_file/0003/267177/SA_Rec_Survey_2013_14.pdf

Glenn, H., Tingley, D., Sánchez Maroño, S., Holm, D., Kell, L., Padda, G., … Kuikka, S. (2012). Trust in the fisheries scientific community. *Marine Policy*, *36*(1), 54–72. https://doi.org/10.1016/j.marpol.2011.03.008

Google. (2017). Google Cloud Machine Learning at Scale | Google Cloud Platform [Web Page]. Retrieved March 3, 2017, from https://cloud.google.com/products/machine-learning/

Google. (2018). Tensorflow detection model zoo [Web Page]. Retrieved May 1, 2018, from https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md

Goudge, H., Morris, E. S., & Sharp, R. (2009). *North Wales Recreational Sea Angler (RSA) Pilot Surveys: Winter Results December 2007 to March 2008* [Report]. Marine Ecological Solutions.

Retrieved from https://1drv.ms/b/s!Au4azAJUYz5JrlYTAgC2lRvCopHl

Goudge, H., Morris, E. S., & Sharp, R. (2010). *North Wales Recreational Sea Angler (RSA) pilot surveys: Summer results July to October 2008* [Report]. Marine Ecological Solutions. Retrieved from https://1drv.ms/b/s!Au4azAJUYz5JrlTfcsiSt6NdmNpn

Green, R., & Giese, M. (2004). Negative Effects of Wildlife Tourism on Wildlife. In K. Higginbottom (Ed.), *Wildlife Tourism. Impacts, Management and Planning* (pp. 81–97). Common Ground Publishing Ltd. Retrieved from https://www.researchgate.net/profile/Noel_Scott/publication/40871766_Wildlife_Tourism_a_strategic_destination_analysis/links/54f6a1c20cf27d8ed71e448e.pdf#page=103

Greenberg, S., & Godin, T. (2015). A tool supporting the extraction of angling effort data from remote camera images. *Fisheries*, *40*(6), 276–287. https://doi.org/10.1080/03632415.2015.1038380

Griffiths, S. P., Bryant, J., Raymond, H. F., & Newcombe, P. A. (2016). Quantifying subjective human dimensions of recreational fishing: Does good health come to those who bait? *Fish and Fisheries*, 1–14. https://doi.org/10.1111/faf.12149

Griffiths, S. P., Sahlqvist, P., Lyle, J., Venables, W., Pollock, K., Sawynok, W., & Moeseneder, C. (2014). *A coordinated national data collection for recreational fishing in Australia* [Report]. Brisbane. https://doi.org/https://doi.org/10.4225/08/584d96f2b4b11

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology*. [Book], John Wiley & Sons.

Guthrie, D. (1991). Creel and angler surveys in fisheries management. In *Proceedings of the International Symposium and Workshop on Creel and Angler Surveys in Fisheries* (p. 528). Houston, Texas: American Fisheries Society.

Hall, G. B., & Close, C. H. (2007). Local knowledge assessment for a small-scale fishery using geographic information systems. *Fisheries Research*, *83*(1), 11–22. https://doi.org/10.1016/j.fishres.2006.08.015

Hamilton, R. J., Giningele, M., Aswani, S., & Ecochard, J. L. (2012). Fishing in the dark-local knowledge, night spearfishing and spawning aggregations in the Western Solomon Islands. *Biological Conservation*, *145*(1), 246–257. https://doi.org/10.1016/j.biocon.2011.11.020

Hannak, A., Sapiezynski, P., Kakhki, A. M., Krishnamurthy, B., Lazer, D., Mislove, A., & Wilson, C. (2013). Measuring personalization of web search. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 527–537). https://doi.org/10.1145/2488388.2488435

Hansen, M. J., Beard, Td., & Hewett, S. W. (2000). Catch rates and catchability of Walleyes in angling and spearing fisheries in Northern Wisconsin lakes. *North American Journal of Fisheries Management*, *20*(1), 109–118. https://doi.org/10.1577/1548-8675

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859).

Hartill, B. W. (2017). *New Zealand Fisheries Assessment Report 2017/43. Review of competition creel survey data provided by the Hawke's Bay Sport Fishing Club* [Report]. Retrieved from https://fs.fish.govt.nz/Page.aspx?pk=113&dk=24485

Hartill, B. W., Payne, G. W., Rush, N., & Bian, R. (2016). Bridging the temporal gap: Continuous and cost-effective monitoring of dynamic recreational fisheries by web cameras and creel surveys. *Fisheries Research*, *183*, 488–497. https://doi.org/10.1016/j.fishres.2016.06.002

Hartley, R., & Zisserman, A. (2004). *Multiple view geometry in computer vision*. [Book], Cambridge University Press.

Harvey, E. S., Fletcher, D., & Shortis, M. R. (2001). A comparison of the precision and accuracy of estimates of reef-fish lengths made by divers and a stereo-video system. *Fishery Bulletin*, *99*(1), 63–71. Retrieved from http://www.geomsoft.com/markss/papers/Harvey_etal2001_comp_precacc.pdf

Harvey, E. S., & Shortis, M. R. (1995). A system for stereo-video measurement of sub-tidal organisms. *Marine Technology Society Journal*, *29*(4), 10–22.

Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., … Di Minin, E. (2018). Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conservation Letters*, *11*(1), e12343. https://doi.org/10.1111/conl.12343

Hayter, A. J. (1984). A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *The Annals of Statistics*, 61–75.

He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2980–2988). Venice, Italy. https://doi.org/10.1109/ICCV.2017.322

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). Retrieved from http://arxiv.org/abs/1512.03385

Heckman, J. J. (1990). Selection bias and self-selection. In *Econometrics* (pp. 201–224). [Book Section], Springer.

Henry, G. W., & Lyle, J. M. (2003). *AGDAFF: The national recreational and indigenous fishing survey. FRDC Project No. 99/158* [Report]. Canberra: Australian Government Department of

Agriculture, Fisheries and Forestry. Retrieved from https://eprints.utas.edu.au/2526/1/Henry_Lyle_Nationalsurvey.pdf

Herfaut, J., Levrel, H., Drogou, M., & et Véron, G. (2010). *Monitoring of recreational fishing of sea-bass (Dicentrarchus labrax) in France: Output from a dual methodology (telephone survey and diary) ICES CM 2010/R:05* [Report]. ICES. Retrieved from http://www.ices.dk/sites/pub/CM Doccuments/CM-2010/R/R0510.pdf

Herfaut, J., Levrel, H., Drogou, M., Thébaud, O., & Véron, G. (2012). *Ecological and economic impacts of marine recreational fishing in France* [Report]. *Working Papers Series*. Amure Publications. Retrieved from https://www.umr-amure.fr/wp-content/uploads/2018/10/D_33_2012.pdf

Hershey, S., Chaudhuri, S., Ellis, D. P. W., Gemmeke, J. F., Jansen, A., Moore, R. C., … Wilson, K. (2017). CNN architectures for large-scale audio classification. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 131–135). IEEE. https://doi.org/10.1109/ICASSP.2017.7952132

Hilborn, R., & Walters, C. J. (1992). *Quantitative fisheries stock assessment: Choice, dynamics and uncertainty*. [Book], Springer Science & Business Media.

Hind, E. J. (2014). Knowledge research : A challenge to established fisheries science. *ICES Journal of Marine Science*, *72*, 341–358. https://doi.org/10.1093/icesjms/fsu169

Hind, E. J. (2015). A review of the past, the present, and the future of fishers’ knowledge research: a challenge to established fisheries science. *ICES Journal of Marine Science*. https://doi.org/10.1093/icesjms/fsu169

Hinsley, A., Lee, T. E., Harrison, J. R., & Roberts, D. L. (2016). Estimating the extent and structure of trade in horticultural orchids via social media. *Conservation Biology*, *30*(5), 1038–1047. https://doi.org/10.1111/cobi.12721

Hipp, R. D., Kennedy, D., & Mistachkin, J. (2018). SQLite. [Computer Program], SQLite Development Team. Retrieved from https://www.sqlite.org/about.html

HMSO. UK Copyright, Designs and Patents Act 1988. (2016). United Kingdom. https://doi.org/10.1049/ir:19880182

Hold, N., Murray, L. G., Pantin, J. R., Haig, J. A., Hinz, H., & Kaiser, M. J. (2015). Video capture of crustacean fisheries data as an alternative to on-board observers. *ICES Journal of Marine Science*, *72*(6), 1811–1821. https://doi.org/10.1093/icesjms/fsv030

Holden, M. J., & Williams, T. (1974). The biology, movements and population dynamics of bass, Dicentrarchus labrax, in English Waters. *Journal of the Marine Biological Association of the*

United Kingdom, *54*, 91–107. https://doi.org/10.1017/S0025315400022098

Holdsworth, J. C., Hartill, B. W., Heinemann, A., & Wynne-Jones, J. (2018). Integrated survey methods to estimate harvest by marine recreational fishers in New Zealand. *Fisheries Research*, *204*(March), 424–432. https://doi.org/10.1016/j.fishres.2018.03.018

Hornik, K. (2016). OpenNLP: Apache OpenNLP Tools Interface. [Computer Program], Apache. Retrieved from https://cran.r-project.org/package=openNLP

Hornik, K., Buchta, C., Hothorn, T., Karatzoglou, A., Meyer, D., & Zeileis, A. (2017). RWeka: R/Weka Interface. [Computer Program]. Retrieved from https://cran.r-project.org/package=RWeka

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., … Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *ArXiv Preprint*, *1704.04861*. Retrieved from http://arxiv.org/abs/1704.04861

Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational propaganda during the UK-EU referendum. *Social Science Research Network*. https://doi.org/10.2139/ssrn.2798311

Howell, D. C. (1997). *Statistical Methods for Psychology* (4th ed.). [Book], Duxbury Press.

Hsiao, Y. H., Chen, C. C., Lin, S. I., & Lin, F. P. (2014). Real-world underwater fish recognition and identification, using sparse representation. *Ecological Informatics*, *23*, 13–21. https://doi.org/10.1016/j.ecoinf.2013.10.002

Hudson, K., & Lee, D. (2010). Biodiversity and tourism: A valuable relationship. *Social Alternatives*, *29*(3), 39.

Hyder, K., Radford, Z., Prellezo, R., Weltersbach, M. S., Lewin, W.-C., Zarauz, L., … Strehlow, H. V. (2017). *Research for the PECH Committee - Marine recreational and semi-subsistence fishing - its value and its impact on fish stock. EURecFish final report for IP/B/PECH/IC/2016-131* [Report]. https://doi.org/10.2861/2728

Hyder, K., Townhill, B., Anderson, L. G., Delany, J., & Pinnegar, J. K. (2015). Can citizen science contribute to the evidence-base that underpins marine policy? *Marine Policy*, *59*(0), 112–120. https://doi.org/10.1016/j.marpol.2015.04.022

Hyder, K., Weltersbach, M. S., Armstrong, M., Ferter, K., Townhill, B., Ahvonen, A., … Strehlow, H. V. (2018). Recreational sea fishing in Europe in a global context – participation rates, fishing effort, expenditure, and implications for monitoring and assessment. *Fish and Fisheries*, *19*(2), 225–243. https://doi.org/10.1111/faf.12251

IBM Corp. (2011). IBM SPSS Statistics for Windows. Released 2011. [Computer Program], Armonk, NY: IBM. Retrieved from https://www.ibm.com/uk-en/products/spss-statistics

ICES. (2012). *Report on the Classification of Stock Assessment Methods developed by SISAM. ICES CM 2012/ACOM/SCICOM:01* [Report]. Retrieved from http://www.ices.dk/community/Documents/SISAM/Report on the Classification of Stock Assessment Methods developed by SISAM.pdf

ICES. (2013a). *ICES Advice for 2014. European sea bass in Divisions IVbc, VIIa, and VIId–h (Irish Sea, Celtic Sea, English Channel, and southern North Sea). Book 5, 5.4.32* [Report]. Retrieved from http://www.ices.dk/sites/pub/Publication Reports/Forms/defaultone.aspx

ICES. (2013b). *ICES WGRFS Report 2013. Report of the Working Group on Recreational Fisheries Surveys (WGRFS)* [Report]. Esporles, Spain. Retrieved from http://www.ices.dk/sites/pub/Publication Reports/Expert Group Report/acom/2013/WGRFS 2013/WGRFS Report 2013.pdf

ICES. (2014a). *ICES Advice for 2015. European sea bass in Divisions IVbc, VIIa, and VIId–h (Irish Sea, Celtic Sea, English Channel, and southern North Sea). Book 5, 5.3.32.* [Report]. Retrieved from http://www.ices.dk/sites/pub/Publication Reports/Forms/defaultone.aspx

ICES. (2014b). *Report of the ICES Advisory Committee 2014. ICES Advice, 2014. Books 1-11.* [Report]. ICES. Retrieved from http://ices.dk/sites/pub/Publication Reports/Forms/defaultone.aspx?RootFolder=%2Fsites%2Fpub%2FPublication Reports%2FICES Advice%2F2014&FolderCTID=0x0120005DAF18EB10DAA049BBB066544D790785

ICES. (2014c). *Report of the Working Group on Recreational Fisheries Surveys (WGRFS), 2014* [Report]. Sukarrieta, Spain: ICES. Retrieved from http://www.ices.dk/sites/pub/Publication Reports/Expert Group Report/acom/2014/WGRFS/WGRFS Final Report 2014.pdf

ICES. (2016). *ICES Advice on fishing opportunities, catch, and effort Celtic Seas and Greater North Sea ecoregions for 2017. Sea bass (Dicentrarchus labrax) in divisions 4.b–c, 7.a, and 7.d–h (central and southern North Sea, Irish Sea, English Channel, Bristol C* [Report]. Retrieved from http://www.ices.dk/sites/pub/Publication Reports/Advice/2016/2016/bss-47.pdf

ICES. (2017a). *ICES Advice on fishing opportunities, catch, and effort. Seabass (Dicentrarchus labrax) in divisions 8.c and 9.a (southern Bay of Biscay and Atlantic Iberian waters)* [Report]. https://doi.org/10.17895/ices.pub.3068

ICES. (2017b). *ICES stock advice. Seabass (Dicentrarchus labrax) in divisions 4.b–c, 7.a, and 7.d–h (central and southern North Sea, Irish Sea, English Channel, Bristol Channel, and Celtic Sea)* [Report]. https://doi.org/10.17895/ices.pub.3334

ICES. (2017c). *ICES WGRFS Report 2017. Report of the Working Group on Recreational Fisheries Surveys (WGRFS) 12-16 June 2017* [Report]. Azores, Portugal. Retrieved from https://www.ices.dk/sites/pub/Publication Reports/Expert Group Report/SSGIEOM/2017/WGRFS/wgrfs_2017.pdf

ICES. (2017d). *Report of the Working Group on Recreational Fisheries Surveys (WGRFS), 6–10 June 2016. ICES CM 2016/SSGIEOM:10* [Report]. Nea Peramos, Greece. Retrieved from https://www.ices.dk/sites/pub/Publication Reports/Expert Group Report/SSGIEOM/2016/WGRFS/WGRFS_2016.pdf

ICES. (2018). *ICES Advice on fishing opportunities, catch, and effort Celtic Seas and Greater North Sea ecoregions. Sea bass (Dicentrarchus labrax) in divisions 4.b–c, 7.a, and 7.d–h for 2019* [Report]. https://doi.org/https://doi.org/10.17895/ices.pub.4472

IMAGENET. (2018). IMAGENET Large Scale Visual Recognition Challenge (ILSVRC) [Web Page]. Retrieved June 6, 2018, from http://www.image-net.org/challenges/LSVRC/

Indecon. (2007). *The Social and Economic Impact to Northern Ireland, and areas within the Loughs Agency, of Recreational Fisheries, Angling and Angling Resources* [Report]. United Kingdom: PWC. Retrieved from https://www.daera-ni.gov.uk/publications/social-and-economic-impact-angling-and-angling-resources-final-report

Inland Fisheries Ireland. (2018). Marine Sport Fish Tagging Programme [Web Page]. Retrieved May 1, 2018, from https://www.fisheriesireland.ie/Tagging/marine-sport-fish-tagging-programme.html

International Game Fish Association. (2018). IGFA Catch Log [Web Page]. Retrieved July 19, 2018, from http://www.igfacatchlog.org/Default.aspx

International Telecommunications Union. (2014). *Measuring the Information Society Report 2014* [Report]. Geneva, Switzerland. Retrieved from http://www.itu.int/en/ITU-D/Statistics/Documents/publications/mis2014/MIS2014_without_Annex_4.pdf

Irish Parliament. Bass (Conservation of Stocks) Regulations 2006. S.I No. 230/2006 (2006). Republic of Ireland.

Isle of Man Government. Bass Fishing Regulations 2016 (2016). Isle of Man.

Jaquet, N. (2006). A simple photogrammetic technique to measure sperm whales at sea. *Marine Mammal Science*, *22*(4), 862–879. https://doi.org/10.1111/j.1748-7692.2006.00060.x

Jentoft, S. (2000). Legitimacy and disappointment in fisheries management. *Marine Policy*, *24*(2), 141–148. https://doi.org/10.1016/S0308-597X(99)00025-1

Jeong, S. J., Yang, Y. S., Lee, K., Kang, J. G., & Lee, D. G. (2013). Vision-based automatic system for non-contact measurement of morphometric characteristics of flatfish. *Journal of Electrical Engineering and Technology, 8*(5), 1194–1201. https://doi.org/10.5370/JEET.2013.8.5.1194

Johannes, R. E., Freeman, M. M. R., & Hamilton, R. J. (2000). Ignore fishers' knowledge and miss the boat. *Fish and Fisheries, 1*(1984), 257–271. https://doi.org/10.1111/j.1467-2979.2000.00019.x

Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., … Barthélémy, D. (2014). Interactive plant identification based on social image data. *Ecological Informatics*, *23*, 22–34. https://doi.org/10.1016/j.ecoinf.2013.07.006

Jones, D. (2014). *Wales Activity Mapping Project Annual Progress Report 2013-14* [Report]. Retrieved from http://www.walesactivitymapping.org.uk/wp-content/uploads/2011/03/Wales-Activity-Mapping-Annual-Progress-Report-2014-Final-LS.pdf

Jones, D. (2017). Wales Activity Mapping Home Page [Web Page]. Retrieved June 15, 2017, from http://www.walesactivitymapping.org.uk

Jones, E., Oliphant, T., Peterson, P., & others. (2017). SciPy: Open source scientific tools for Python. [Computer Program]. Retrieved from http://www.scipy.org/

Jones, K. (2009). *South Australian Recreational Fishing Survey 2007/08. Paper No. 54* [Report]. Adelaide: Government of South Australia. Retrieved from http://fish.gov.au/reports/Documents/Jones_2009.pdf

Jung, A. (2018). imgaug: Image augmentation for machine learning experiments. [Computer Program]. Retrieved from https://github.com/aleju/imgaug

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing* (Vol. 2). [Book], Pearson.

Kafas, A., Mclay, A., Chimienti, M., Scott, B. E., Davies, I., & Gubbins, M. (2017). ScotMap: Participatory mapping of inshore fishing activity to inform marine spatial planning in Scotland. *Marine Policy, 79*(January), 8–18. https://doi.org/10.1016/j.marpol.2017.01.009

Kaplan, I. M., & McCay, B. J. (2004). Cooperative research, co-management and the social dimension of fisheries science and management. *Marine Policy*, *28*(3), 257–258. https://doi.org/10.1016/j.marpol.2003.08.003

Katz, M. L., & Shapiro, C. (1994). Systems competition and network effects. *The Journal of Economic Perspectives*, *8*, 93–115. https://doi.org/10.2307/2138538

Kazemi, V., & Sullivan, J. (2014). One millisecond face alignment with an assemble of regression trees. In *27th IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1867–1874). Columbus: IEEE Computer Society. https://doi.org/10.13140/2.1.1212.2243

Keller, K., Steffe, A. S., Lowry, M., Murphy, J. J., & Suthers, I. M. (2016). Monitoring boat-based recreational fishing effort at a nearshore artificial reef with a shore-based camera. *Fisheries Research*, *181*, 84–92. https://doi.org/10.1016/j.fishres.2016.03.025

Kelley, D. F. (1987). Food of bass in U.K. waters. *J. Mar. Biol. Ass. U.K.*, *67*(1987), 275–286.

Kenter, J. O., Bryce, R., Davies, A., Jobstvogt, N., Watson, V., Ranger, S., … Reed, M. S. (2013). *The value of potential marine protected areas in the UK to divers and sea anglers.* [Report]. Cambridge; UK.: UNEP-WCMC. Retrieved from http://uknea.unep-wcmc.org/LinkClick.aspx?fileticket=Mb8nUAphh%2BY%3D&tabid=82

King, D. E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, *10*, 1755–1758.

Kish, L. (1995). *Survey sampling*. [Book], Wiley-Blackwell.

Konovalov, D. A., Domingos, J. A., Bajema, C., White, R. D., & Jerry, D. R. (2017). Ruler Detection for Automatic Scaling of Fish Images. In *Proceedings of the International Conference on Advances in Image Processing* (pp. 90–95). New York, NY, USA: ACM. https://doi.org/10.1145/3133264.3133271

Koster, M. (1994). The Web Robots Pages [Web Page]. Retrieved July 19, 2017, from http://www.robotstxt.org/orig.html

Kraan, M., Uhlmann, S., Steenbergen, J., Helmond, A. T. M. Van, Van Helmond, A. T. M., & Van Hoof, L. (2013). The optimal process of self-sampling in fisheries: Lessons learned in the Netherlands. *Journal of Fish Biology*, *83*(4), 963–973. https://doi.org/10.1111/jfb.12192

Kühl, H. S., & Burghardt, T. (2013). Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in Ecology & Evolution*, *28*(7), 432–441. https://doi.org/10.1016/J.TREE.2013.02.013

Kumar, N., Belhumeur, P. N., Biswas, A., Jacobs, D. W., Kress, W. J., Lopez, I. C., & Soares, V. B. (2012). A Computer Vision System for Automatic Plant Species Identification What Plant Species is this ? In *European Conference on Computer Vision* (pp. 1–14). Springer-Verlag. https://doi.org/10.1007/978-3-642-33709-3_36

Ladle, R. J., Correia, R. A., Do, Y., Joo, G. J., Roberge, J. M., Jepson, P., … Jepson, P. (2016). Conservation culturomics. *Frontiers in Ecology and the Environment*, *14*(5), 269–275. https://doi.org/10.1002/fee.1260

Laist, D. W. (1997). Impacts of Marine Debris: Entanglement of marine life in marine debris including a comprehensive list of species with entanglement and ingestion records. In J. M. Coe & D. B. Rogers (Eds.), *Marine Debris*. New York, New York, USA: Springer. https://doi.org/https://doi.org/10.1007/978-1-4613-8486-1_10

Lancaster, D., Dearden, P., Haggarty, D. R., Volpe, J. P., & Ban, N. C. (2017). Effectiveness of shore-based remote camera monitoring for quantifying recreational fisher compliance in marine conservation areas. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(4), 804–813. https://doi.org/10.1002/aqc.2736

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159. https://doi.org/10.2307/2529310

Lavrakas, P. (2008). Self-Selection Bias. In P. Lavrakas (Ed.), *Encyclopedia of Survey Research Methods*. [Book Section], Thousand Oaks, California: Sage Publications. https://doi.org/10.4135/9781412963947

Lawrence, K. S., & Spurgeon, J. (2007). *Economic evaluation of inland fisheries: Welfare benefits of inland fisheries in England and Wales. Science Report SC050026/SR1* [Report]. *Environment Agency*. Bristol, UK: Environment Agency. Retrieved from https://www.gov.uk/government/publications/economic-evaluation-of-inland-fisheries

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, *343*(March), 1203–1205. https://doi.org/10.1126/science.1248506

Lee, D.-J., Schoenberger, R. B., Shiozawa, D., Xu, X., & Zhan, P. (2004). Contour matching for a fish recognition and migration-monitoring system. In K. G. Harding (Ed.), *SPIE Proceedings 5606, Two- and Three-Dimensional Vision Systems for Inspection, Control, and Metrology II* (p. 12). https://doi.org/10.1117/12.571789

Léopold, M., Guillemot, N., Rocklin, D., & Chen, C. (2014). A framework for mapping small-scale coastal fisheries using fishers' knowledge. *ICES Journal of Marine Science*, *71*(7), 1781–1792. https://doi.org/10.1093/icesjms/fst204

Lepczyk, C. A. (2005). Integrating published data and citizen science to describe bird diversity across a landscape. *Journal of Applied Ecology*, *42*(4), 672–677. https://doi.org/10.1111/j.1365-2664.2005.01059.x

Lerman, K. (2007). User Participation in Social Media: Digg Study. In *IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology Workshops* (pp. 255–258). arXiv. Retrieved from https://arxiv.org/abs/0708.2414

Lewin, W.-C., Arlinghaus, R., & Mehner, T. (2006). Documented and potential biological impacts of recreational fishing: Insights for management and conservation. *Reviews in Fisheries Science*, *14*(4), 305–367. https://doi.org/10.1080/10641260600886455

Lewin, W.-C., Strehlow, H. V., Ferter, K., Hyder, K., Niemax, J., Herrmann, J.-P., & Weltersbach, M. S. (2018). Estimating post-release-mortality of the European sea bass based on experimental angling. *ICES Journal of Marine Science*, *75*(4), 1483–1495. https://doi.org/https://doi.org/10.1093/icesjms/fsx240

Lin, T., RoyChowdhury, A., & Maji, S. (2015). Bilinear CNN Models for Fine-grained Visual Recognition. In *IEEE International Conference on Computer Vision* (pp. 1–14). Santiago: IEEE. https://doi.org/https://doi.org/10.1109/ICCV.2015.170

Linke, S., & Bruckmeier, K. (2015). Co-management in fisheries - Experiences and changing approaches in Europe. *Ocean and Coastal Management*, *104*, 170–181. https://doi.org/10.1016/j.ocecoaman.2014.11.017

Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. In *Eighth IEEE International Conference on Data Mining* (pp. 413–422). IEEE Computer Society. https://doi.org/http://doi.ieeecomputersociety.org/10.1109/ICDM.2008.17

Lovell, S. J., Steinback, S., & Hilger, H. (2013). *The Economic Contribution of Marine Angler Expenditures in the United States, 2011* [Report]. *NOAA Technical Memorandum* (Vol. NMFS-F/SPO). U.S. Department of Commerce. Retrieved from http://unicat.bangor.ac.uk:4550/resserv?genre=report&issn=&title=Economic+Contribution+of+Marine+Angler+Expenditures+in+the+United+States%252C+2006+%2528Revised+December+2008%2529&volume=&issue=&date=2008-01-01&atitle=&spage=321&aulast=Gentner&sid=ProQ:Aq

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

Macdonald, P., Angus, C. H., Cleasby, I. R., & Marshall, C. T. (2014). Fishers' knowledge as an indicator of spatial and temporal trends in abundance of commercial fish species: Megrim (*Lepidorhombus whiffiagonis*) in the northern North Sea. *Marine Policy*, *45*, 228–239. https://doi.org/10.1016/j.marpol.2013.11.001

Maggs, J. Q., Mann, B. Q., Potts, W. M., & Dunlop, S. W. (2016). Traditional management strategies fail to arrest a decline in the catch-per-unit-effort of an iconic marine recreational fishery species with evidence of hyperstability. *Fisheries Management and Ecology*, *23*(3–4), 187–199. https://doi.org/10.1111/fme.12125

Maine Department of Marine Resources. (2016). 2016 Maine Striped Bass Regulations [Web Page]. Retrieved January 10, 2018, from http://www.fishingnortheast.net/wp-content/uploads/2016/10/2016striped_bass_regs.pdf

MaPP. (2016). MaPP | Marine Plan Partnership for the North Pacific Coast [Web Page]. Retrieved December 13, 2017, from http://mappocean.org/

Marchal, P., Andersen, B., Caillart, B., Eigaard, O., Guyader, O., Hovgaard, H., … Santurtún, M. (2007). Impact of technological creep on fishing effort and fishing mortality, for a selection of European fleets. *ICES Journal of Marine Science*, *64*(1), 192–209. https://doi.org/10.1093/icesjms/fsl014

Markham, A., & Buchanan, E. (2012). *Ethical Decision-Making and Internet Research: Recommendations from the AoIR Ethics Working Committee* [Report]. Retrieved from http://www.aoir.org/documents/ethics-guide

Martin, D. R., Chizinski, C. J., Eskridge, K. M., & Pope, K. L. (2014). Using posts to an online social network to assess fishing effort. *Fisheries Research*, *157*(0), 24–27. https://doi.org/http://dx.doi.org/10.1016/j.fishres.2014.03.013

Martin, D. R., Pracheil, B. M., Deboer, J. A., Wilde, G. R., & Pope, K. L. (2012). Using the internet to understand angler behavior in the information age. *Fisheries*, *37*(10), 458–463. https://doi.org/10.1080/03632415.2012.722875

Maurstad, A. (2002). Fishing in murky waters-ethics and politics of research on fisher knowledge. *Marine Policy*, *26*(3), 159–166. https://doi.org/10.1016/S0308-597X(01)00045-8

Mawle, G. W., & Peirson, G. (2009). *Economic evaluation of inland fisheries: managers report from science project SC050026/SR2. 58.* [Report]. Bristol, UK: Environment Agency. Retrieved from https://www.gov.uk/government/publications/economic-evaluation-of-inland-fisheries

Mbaiwa, J. E. (2003). The socio-economic and environmental impacts of tourism development on the Okavango Delta, north-western Botswana. *Journal of Arid Environments*, *54*(2), 447–467. https://doi.org/10.1006/jare.2002.1101

McClanahan, T. R. (1994). Kenyan coral reef lagoon fish: effects of fishing, substrate complexity, and sea urchins. *Coral Reefs*, *13*(4), 231–241. https://doi.org/https://doi.org/10.1007/BF00303637

McClenachan, L. (2009). Historical declines of goliath grouper populations in South Florida, USA. *Endangered Species Research*, *7*(3), 175–181. https://doi.org/10.3354/esr00167

McGlennon, D., & Kinloch, M. A. (1997). Evaluation of the bus-route creel survey method in a large Australian marine recreational fishery - II. Pilot surveys and optimal sampling allocation. *Fisheries Research*, *33*(1–3), 89–99. https://doi.org/10.1016/S0165-7836(97)00067-2

McManus, A., Storey, J., & White, J. (2011). *Identifying the health and well-being benefits of recreational fishing, Report No. 2011/217* [Report]. Fisheries Research and Development Corporation and the Centre of Excellence for Science, Seafood and Health; Curtin Health Innovation Research Institute. Retrieved from http://www.tarfish.org/documents/Health and Wellbeing of Rec Fishing 2012.pdf

McMinn, C. (2013). *Report on the Survey of Recreational Sea Angling in Northern Ireland* [Report]. Belfast: Agri-Food and Biosciences Institute. Retrieved from https://www.daera-ni.gov.uk/sites/default/files/publications/dard/report-on-the-survey-of-recreational-sea-angling-in-northern-ireland.pdf

McPhee, D. P., Leadbitter, D., & Skilleter, G. A. (2002). Swallowing the bait: is recreational fishing in Australia ecologically sustainable? *Pacific Conservation Biology*, *8*, 40–51.

Microsoft. (2008). SQL Server 2008 Standard Edition. [Computer Program], Redmond, USA: Microsoft.

Microsoft. (2017). Machine learning | Microsoft Azure [Web Page]. Retrieved March 3, 2017, from https://azure.microsoft.com/en-gb/services/machine-learning/

Ministry for Primary Industries. (2012). *Report from the Fisheries Assessment Plenary, May 2012. Stock assessments and yield estimates* [Report]. Wellington, New Zealand. Retrieved from https://www.mpi.govt.nz/dmsdocument/4185/send

Minteer, B. A., & Collins, J. P. (2005). Why we need an “ecological ethics.” *Frontiers in Ecology and the Environment*, *3*(6), 332–337. Retrieved from http://www.jstor.org/stable/3868567

Miranda, J. M., & Romero, M. (2017). A prototype to measure rainbow trout’s length using image processing. *Aquacultural Engineering*, *76*, 41–49. https://doi.org/10.1016/J.AQUAENG.2017.01.003

Monkman, G. G., Cambie, G., Hyder, K., Armstrong, M., Roberts, A., & Kaiser, M. J. (2015). *Socioeconomic and Spatial Review of Recreational Sea Angling in Wales* [Report]. Bangor, North Wales: Bangor University, Fisheries and Conservation Science Group. Retrieved from http://fisheries-conservation.bangor.ac.uk/documents/52.pdf

Monkman, G. G., Hyder, K., Kaiser, M. J., & Vidal, F. P. (2018). Accurate estimation of fish length in single camera photogrammetry with a fiducial marker. *Manuscript in Review*.

Monkman, G. G., Kaiser, M., & Hyder, K. (2018). The ethics of using social media in fisheries research. *Reviews in Fisheries Science & Aquaculture*, *26*(2), 235–242. https://doi.org/10.1080/23308249.2017.1389854

Mori, E., Di Bari, P., & Coraglia, M. (2018). Interference between roe deer and Northern chamois in the Italian Alps: are Facebook groups effective data sources? *Ethology Ecology & Evolution*, *30*(3), 277–284. https://doi.org/10.1080/03949370.2017.1354922

Muir, A. M., Vecsei, P., & Krueger, C. C. (2012). A perspective on perspectives: Methods to reduce variation in shape analysis of digital images. *Transactions of the American Fisheries Society*, *141*(4), 1161–1170. https://doi.org/10.1080/00028487.2012.685823

Murdock, S. H., Loomis, D. K., Ditton, R. B., & Hoque, N. (1996). The implications of demographic change for recreational fisheries management in the United States. *Human Dimensions of Wildlife, 1*(4), 14–37. https://doi.org/10.1080/10871209609359076

Murphy, K. P. (2012). *Machine learning : A probabilistic perspective*. [Book], MIT Press.

Naidoo, R., Weaver, L. C., Diggle, R. W., Matongo, G., Stuart-Hill, G., & Thouless, C. (2016). Complementary benefits of tourism and hunting to communal conservancies in Namibia. *Conservation Biology*, *30*(3), 628–638. https://doi.org/10.1111/cobi.12643

Narayanan, A., & Shmatikov, V. (2009). De-anonymizing social networks. In *Security and Privacy, 2009 30th IEEE Symposium on* (pp. 173–187).

Nardi, B. (2015). When fieldnotes seem to write themselves: ethnography online. In R. Sanjek & S. W. Tratner (Eds.), *EFieldNotes: The Makings of Anthropology in the Digital World* (p. 192). [Book], University of Pennsylvania Press.

National Oceanic and Atmospheric Administration. Magnuson-Stevens Fishery Conservation and Management Act - 2007 Reauthorization Implentation (2007).

National Oceanic and Atmospheric Administration. (2015a). *A Cost Comparison of At-Sea Observers and Electronic Monitoring for a Hypothetical Midwater Trawl Herring / Mackerel Fishery*. [Report]. Retrieved from https://www.greateratlantic.fisheries.noaa.gov/fish/em_cost_assessment_for_gar_herring_150904_v6.pdf

National Oceanic and Atmospheric Administration. (2015b). *A Preliminary Cost Comparison of At-sea Monitoring and Electronic Monitoring for a Hypothetical Groundfish Sector* [Report]. Retrieved from https://www.greateratlantic.fisheries.noaa.gov/stories/2015/september/em_cost_assessment_for_gar_multispecies_2015_06_10.pdf

National Oceanic and Atmospheric Administration. (2016). NOAA Office of Science and Technology National Marine Fisheries Service: MRIP Survey Data for 2016 [Web Page]. Retrieved November 14, 2017, from http://www.st.nmfs.noaa.gov/recreational-fisheries/data-and-documentation/downloads

National Oceanic and Atmospheric Administration. (2017). Marine Recreational Information Program [Web Page]. Retrieved December 13, 2017, from https://www.st.nmfs.noaa.gov/recreational-fisheries/index

National Research Council. (2006). *Committee on the Review of Recreational Fisheries Survey Methods: Review of recreational fisheries survey methods.* [Report]. Washington D.C.: The

National Academies Press. https://doi.org/doi.org/10.17226/11616

Nautilus Consultants. (2000). *Study into Inland and Sea Fisheries in Wales: Final Report* [Report]. Nautilus Consultants Ltd. Retrieved from http://www.nautilus-consultants.co.uk/sites/default/files/wales.pdf

Needle, C. L., Dinsdale, R., Buch, T. B., Catarino, R. M. D., Drewery, J., & Butler, N. (2015). Scottish science applications of Remote Electronic Monitoring. *ICES Journal of Marine Science*, *72*(4), 1214–1229. https://doi.org/https://doi.org/10.1093/icesjms/fsu225

NERC. (1970). *National Angling Survey* [Report]. NERC. Retrieved from Unavailable Online

Nery, M. S., Machado, A. M., Campos, M. F. M., Padua, F. L. C., Carceroni, R., & Queiroz-Neto, J. P. (2005). Determining the Appropriate Feature Set for Effective Fish Classification Tasks. In *Brazilian Symposium on Computer Graphics and Image Processing* (pp. 173–180).

netmarketshare.com. (2017). Search engine market share [Web Page]. Retrieved February 16, 2017, from https://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4&qpcustomd=0

Neuswanger, J. R., Wipfli, M. S., & Rosenberger, A. E. (2016). Measuring fish and their physical habitats : Versatile 2-D and 3-D video techniques with user-friendly software. *Canadian Journal of Fisheries and Aquatic Sciences*, *13*(June), 1–48. https://doi.org/10.1139/cjfas-2016-0010

New Zealand Parliament. Privacy Act 1993. No 28. (1993). Retrieved from http://www.legislation.govt.nz/act/public/1993/0028/latest/DLM296639.html

New Zealand Parliament. Copyright Act 1994. No 143 section 43 - Research or private study. (1994).

Newing, H., Eagle, C. M., Puri, R. K., & Watson, C. W. (2010). *Conducting Research in Conservation: Social Science Methods and Practice*. [Book], Routledge. https://doi.org/10.4324/9780203846452

Nguyen, T. X., Winger, P. D., Legge, G., Dawe, E. G., & Mullowney, D. R. (2014). Underwater observations of the behaviour of snow crab (*Chionoecetes opilio*) encountering a shrimp trawl off northeast Newfoundland. *Fisheries Research*, *156*, 9–13. https://doi.org/10.1016/J.FISHRES.2014.04.013

Nielsen, J. (2017). Participation Inequality: The 90-9-1 Rule for Social Features [Web Page]. Retrieved November 30, 2017, from https://www.nngroup.com/articles/participation-inequality/

Nilsback, M. E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. In *6th Indian Conference on Computer Vision, Graphics and Image Processing, ICVGIP 2008* (pp. 722–729). https://doi.org/10.1109/ICVGIP.2008.47

Norval, C., & Henderson, T. (2017). Contextual consent: Ethical mining of social media for health research. *ArXiv*. Retrieved from http://arxiv.org/abs/1701.07765

NVIDIA. (2018). NVIDIA Accelerated Computing, CUDA Zone [Web Page]. Retrieved June 1, 2018, from https://developer.nvidia.com/cuda-zone

O'Toole, A. C., Hanson, K. C., & Cooke, S. J. (2009). The effect of shoreline recreational angling activities on aquatic and riparian habitat within an urban environment: implications for conservation and management. *Environmental Management*, *44*(2), 324–334.

Odiase, J. I., & Ogbonmwan, S. M. (2007). Correlation analysis: Exact permutation paradigm. *Matematički Vesnik*, *59*(4), 161–170.

Office for Human Research Protections. Department of Health and Human Services. Code of Federal Regulations: Title 45, Part 46. Protection of Human Subjects (2009). USA. Retrieved from https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html

Office for National Statistics. (2018). Consumer price inflation time series (MM23). CPI MONTHLY RATE 09.3.2 : EQUIP FOR SPORT CAMPING & OPEN-AIR RECREATION. Retrieved June 20, 2018, from https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/d7n8

Olsen, B., & Thuen, T. (2013). Secret places: On the management of knowledge and information about landscape and yields in northern Norway. *Human Ecology*, *41*(2), 273–283. Retrieved from http://www.jstor.org/stable/24015698

OpenCV team. (2018). OpenCV: Camera Calibration and 3D Reconstruction [Web Page]. Retrieved April 23, 2018, from https://docs.opencv.org/master/d9/d0c/group__calib3d.html

Ordnance Survey. (2015). 1:50 000 Scale Gazetteer | Business and government | Ordnance Survey. Retrieved May 10, 2015, from http://www.ordnancesurvey.co.uk/business-and-government/products/50k-gazetteer.html

Ormsby, J. (2004). *A Review of the Social, Motivational and Experiential Characteristics of Recreational Anglers from Queensland and the Great Barrier Reef Region.* Townsville, Australia: Great Barrier Reef Marine Park Authority. Retrieved from http://elibrary.gbrmpa.gov.au/jspui/handle/11017/384

Outwit Technologies. (2013). Outwit Hub Professional. [Computer Program], Outwit Technologies. Retrieved from https://www.outwit.com/products/hub/

Ouyang, W., Zhou, H., Li, H., Li, Q., Yan, J., & Wang, X. (2018). Jointly learning deep features, deformable parts, occlusion and classification for pedestrian detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(8), 1874–1887. https://doi.org/10.1109/TPAMI.2017.2738645

Oxford Dictionaries. (2018). midnight | Synonyms of midnight by Oxford Dictionaries Thesaurus. Retrieved June 15, 2018, from https://en.oxforddictionaries.com/thesaurus/midnight

Page, K. (2012). The four principles: Can they be measured and do they predict ethical decision making? *BMC Medical Ethics*, *13*(1), 10. https://doi.org/10.1186/1472-6939-13-10

Palomino, M., Taylor, T., Göker, A., Isaacs, J., & Warber, S. (2016). The online dissemination of nature–health concepts: Lessons from sentiment analysis of social media relating to “nature-deficit disorder.” *International Journal of Environmental Research and Public Health*, *13*(1), 142. https://doi.org/10.3390/ijerph13010142

Papenfuss, J. T., Phelps, N., Fulton, D., & Venturelli, P. A. (2015). Smartphones reveal angler behavior: A case study of a popular mobile fishing application in Alberta, Canada. *Fisheries*, *40*(7), 318–327. https://doi.org/10.1080/03632415.2015.1049693

Parnell, E. P., Dayton, P. K., Fisher, R. A., Loarie, C. C., Darrow, R. D., Parnell, P. E., … Darrow, R. D. (2010). Spatial patterns of fishing effort off San Diego: implications for zonal management and ecosystem function. *Ecological Applications*, *20*(8), 2203–2222. https://doi.org/10.1890/09-1543.1

Pasco, G., Whittaker, C., Elliot, S., & Swarbrick, J. (2009). *Northern Irish CCTV Trials: 2009* [Report]. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.177.5899&rep=rep1&type=pdf

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). ENet: A deep neural network architecture for real-time semantic segmentation. *ArXiv Preprint*. Retrieved from http://arxiv.org/abs/1606.02147

Pauly, D. (1995). Anecdotes and the shifting baseline syndrome of fisheries. *Trends in Ecology & Evolution*, *10*(10), 430. https://doi.org/10.1016/S0169-5347(00)89171-5

Pauly, D. (1998). *Rationale for reconstructing catch time series* [Report]. *European Commission Fisheries Cooperation Bulletin* (Vol. 11). Retrieved from Unavailable Online

Pauly, D., & Morgan, G. R. (1987). Length-Based Methods in Fisheries Research. In *The Theory and Application of Length-Based Stock Assessment Methods* (pp. 1–459). Mazara del Vallo, Sicily.

Pawson, M. G., Glenn, H., & Padda, G. (2008). The definition of marine recreational fishing in Europe. *Marine Policy*, *32*(3), 339–350. https://doi.org/10.1016/j.marpol.2007.07.001

Pawson, M. G., Kelley, D. F., & Pickett, G. D. (1987). The distribution and migrations of bass, Dicentrarchus labrax L., in waters around England and Wales as shown by tagging. *Journal of the Marine Biological Association of the United Kingdom*, *67*, 183–217. https://doi.org/10.1017/S0025315400026448

Pawson, M. G., & Pickett, G. D. (1987). *The bass (*Dicentrarchus labrax*) and management of its fishery in England and Wales*. *Great Britain Ministry of Agriculture Fisheries and Food*

Laboratory Leaflet. Lowestoft. Retrieved from https://www.cefas.co.uk/publications/lableaflets/lableaflet59.pdf

Pawson, M. G., & Pickett, G. D. (1996). The annual pattern of condition and maturity in bass, *Dicentrarchus labrax*, in waters around England and Wales. *Journal of the Marine Biological Association of the United Kingdom*, *76*(1), 107–125.

Pawson, M. G., Pickett, G. D., Leballeur, J., Brown, M., & Fritsch, M. (2007a). Migrations, fishery interactions, and management units of sea bass (Dicentrarchus labrax) in northwest Europe. *ICES Journal of Marine Science*, *64*(2), 332–345. https://doi.org/10.1093/icesjms/fsl035

Pawson, M. G., Pickett, G. D., Leballeur, J., Brown, M., & Fritsch, M. (2007b). Migrations, fishery interactions, and management units of sea bass (Dicentrarchus labrax) in Northwest Europe. *ICES Journal of Marine Science*, *64*(2), 332–345. https://doi.org/10.1093/icesjms/fsl035

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *ArXiv Preprint*, 8. Retrieved from http://arxiv.org/abs/1712.04621

Person-Le Ruyet, J., Mahé, K., Le Bayon, N., & Le Delliou, H. (2004). Effects of temperature on growth and metabolism in a Mediterranean population of European sea bass, *Dicentrarchus labrax*. *Aquaculture*, *237*(1–4), 269–280. https://doi.org/10.1016/J.AQUACULTURE.2004.04.021

Perzia, P., Battaglia, P., Consoli, P., Andaloro, F., & Romeo, T. (2016). Swordfish monitoring by a GIS-based spatial and temporal distribution analysis on harpoon fishery data: A case of study in the central Mediterranean Sea. *Fisheries Research*, *183*, 424–434. https://doi.org/10.1016/j.fishres.2016.07.006

Pew Research Center. (2018). Smartphone ownership on the rise in emerging economies [Web Page]. Retrieved August 10, 2018, from http://www.pewglobal.org/2018/06/19/2-smartphone-ownership-on-the-rise-in-emerging-economies/

Pickett, G. D. (1990). *Assessment of the UK bass fishery using a log-book-based catch recording system. Fisheries Research Technical Report No. 90* [Report]. Retrieved from https://www.cefas.co.uk/publications/techrep/tech90.pdf

Poli, B. M., Parisi, G., Zampacavallo, G., Mecatti, M., Lupi, P., Gualtieri, M., & Franci, O. (2001). Quality outline of European sea bass (*Dicentrarchus labrax*) reared in Italy: Shelf life, edible yield, nutritional and dietetic traits. *Aquaculture*, *202*(3–4), 303–315.

https://doi.org/10.1016/S0044-8486(01)00780-3

Pollock, K. H., Hoenig, J. M., Jones, C. M., Robson, D. S., & Greene, C. J. (1997). Catch rate estimation for roving and access point surveys. *American Journal of Fisheries Management*, *17*, 11–19. https://doi.org/10.1577/1548-8675(1997)017<0011:CREFRA>2.3.CO;2

Pollock, K. H., Jones, C. M., & Brown, T. L. (1994). Angler survey methods and their applications in fisheries management. *American Fisheries Society Special Publication*, *25*, i–xvii, 1-371.

Pomeroy, R., & Douvere, F. (2008). The engagement of stakeholders in the marine spatial planning process. *Marine Policy*, *32*(5), 816–822. https://doi.org/10.1016/j.marpol.2008.03.017

Post, J. R., Sullivan, M., Cox, S., Lester, N. P., Walters, C. J., Parkinson, E. A., … Shuter, B. J. (2002). Canada's recreational fisheries: The invisible collapse? *Fisheries*, *27*(1), 6–17. https://doi.org/10.1577/1548-8446(2002)027<0006:CRF>2.0.CO;2

Poushter, J. (2016). *Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies* [Report]. *Pew Research Center*. Pew Research Center. https://doi.org/10.1017/CBO9781107415324.004

Pretty, J., Peacock, J., & Hine, R. (2007). Green exercise in the UK countryside: Effects on health and psychological well-being, and implications for policy and planning. *Journal of Environmental Planning and Management*, (50), 211–231. https://doi.org/10.1080/09640560601156466

Princeton University. (2018). WordNet | A Lexical Database for English [Web Page]. Retrieved June 15, 2018, from https://wordnet.princeton.edu/

Pullin, A. S., & Knight, T. M. (2003). Support for decision making in conservation practice: An evidence-based approach. *Journal for Nature Conservation*, *11*(2), 83–90. https://doi.org/10.1078/1617-1381-00040

Pullin, A. S., & Stewart, G. B. (2006). Guidelines for systematic review in conservation and environmental management. *Conservation Biology*, *20*(6), 1647–1656. https://doi.org/10.1111/j.1523-1739.2006.00485.x

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Computer Program]. Retrieved from http://www.r-project.org/

Radford, A., Riddington, G., & Gibson, H. (2007). *Economic evaluation of inland fisheries: The economic impact of freshwater angling in England & Wales. Science Report – SC050026/SR2* [Report]. Environment Agency. Retrieved from https://www.gov.uk/government/publications/economic-evaluation-of-inland-fisheries

Radford, Z., Hyder, K., Mugerza, E., Ferter, K., Prellezo, R., Townhill, B., … Weltersbach, M. S.

(2018). The impact of marine recreational fishing on key fish stocks in European waters. *PloS One*, *13*(9). https://doi.org/https://doi.org/10.1371/journal.pone.0201666

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031

Ricard, D., Minto, C., Jensen, O. P., & Baum, J. K. (2012). Examining the knowledge base and status of commercially exploited marine species with the RAM Legacy Stock Assessment Database. *Fish and Fisheries*, *13*(4), 380–398. https://doi.org/10.1111/j.1467-2979.2011.00435.x

Richards, D. R., & Friess, D. A. (2015). A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs. *Ecological Indicators*, *53*, 187–195. https://doi.org/10.1016/j.ecolind.2015.01.034

Richardson, E. A. (2006). *Socioeconomic and ecological implications of an ecosystem approach to marine resource management for Wales, UK* [Thesis]. University of Wales, Bangor.

Richardson, E. A., Kaiser, M. J., Edwards-Jones, G., & Ramsay, K. (2006). Trends in sea anglers catches of trophy fish in relation to stock size. *Fisheries Research (Amsterdam)*, *82*(1–3), 253–262. https://doi.org/http://0-dx.doi.org.unicat.bangor.ac.uk/10.1016/j.fishres.2006.05.014

Rizgalla, J., Shinn, A. P., Ferguson, H. W., Paladini, G., Jayasuriya, N. S., & Bron, J. E. (2017). A novel use of social media to evaluate the occurrence of skin lesions affecting wild dusky grouper, *Epinephelus marginatus* (Lowe, 1834), in Libyan coastal waters. *Journal of Fish Diseases*, *40*(5), 609–620. https://doi.org/10.1111/jfd.12540

Roberts, A., Munday, M., Roche, N., Brown, A., Armstrong, M., Hargreaves, J., … Hyder, K. (2017). Assessing the contribution of recreational sea angling to the English economy. *Marine Policy*, *83*, 146–152. https://doi.org/https://doi.org/10.1016/j.marpol.2017.05.028

Rocklin, D., Levrel, H., Drogou, M., Herfaut, J., & Veron, G. (2014). Combining telephone surveys and fishing catches self-report: The French sea bass recreational fishery assessment. *PloS One*, *9*(1), e87271. https://doi.org/10.1371/journal.pone.0087271

Rodgers, W. L., Brown, C., & Duncan, G. J. (1993). Errors in survey reports of earnings, hours worked, and hourly wages. *Journal of the American Statistical Association*, *88*(424), 1208–1218. https://doi.org/10.2307/2291259

Rogers, T. D., Cambiè, G., & Kaiser, M. J. (2017). Determination of size, sex and maturity stage of free swimming catsharks using laser photogrammetry. *Marine Biology*, *164*(11), 1–11. https://doi.org/10.1007/s00227-017-3241-7

Rohner, C. A., Richardson, A. J., Prebble, C. E. M., Marshall, A. D., Bennett, M. B., Weeks, S. J., …

Pierce, S. J. (2015). Laser photogrammetry improves size and demographic estimates for whale sharks. *PeerJ*, *3:e886*. https://doi.org/10.7717/peerj.886

Rosen, S., Jörgensen, T., Hammersland-White, D., & Holst, J. C. (2013). DeepVision: a stereo camera system provides highly accurate counts and lengths of fish passing inside a trawl. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(10), 1456–1467. https://doi.org/10.1139/cjfas-2013-0124

Rosner, D. (2017). Health, climate change, and the descent of science-based policy. *The Milbank Quarterly*, *95*(1), 36–39. https://doi.org/10.1111/1468-0009.12243

Ryan, K. L., Hall, N. G., Lai, E. K., Smallwood, C. B., Taylor, S. M., & Wise, B. S. (2015). *Fisheries Research Report No. 268. State-wide survey of boat-based recreational fishing in Western Australia 2013/14* [Report]. Retrieved from http://www.fish.wa.gov.au/Documents/research_reports/frr268.pdf

Rycroft, N., Radcliffe, K., & Atema, J. (2013). A photographic method for lobster morphometry aimed at site discrimination. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(11), 1658–1665. https://doi.org/10.1139/cjfas-2013-0103

Saillant, E., Fostier, A., Menu, B., Haffray, P., & Chatain, B. (2001). Sexual growth dimorphism in sea bass *Dicentrarchus labrax*. *Aquaculture*, *202*(3–4), 371–387. https://doi.org/10.1016/S0044-8486(01)00786-4

Scaramuzza, D., Martinelli, A., & Siegwart, R. (2006). A toolbox for easily calibrating omnidirectional cameras. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 5695–5701).

Scheyvens, R. (1999). Ecotourism and the empowerment of local communities. *Tourism Management*, *20*(2), 245–249. https://doi.org/10.1016/S0261-5177(98)00069-7

Schiller, L., Alava, J. J., Grove, J., Reck, G., & Pauly, D. (2013). *A Reconstruction of Fisheries Catches for the Galapagos Islands, 1950-2010* [Report]. Retrieved from https://www.semanticscholar.org/paper/A-Reconstruction-of-Fisheries-Catches-for-the-%2C-Schillera-Alavaa/9cedbc7bd87a2cea4e5dc2b77d2d4df24f597d47

Schiller, L., Alava, J. J., Grove, J., Reck, G., & Pauly, D. (2015). The demise of Darwin's fishes: Evidence of fishing down and illegal shark finning in the Galpagos Islands. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *25*(3), 431–446. https://doi.org/10.1002/aqc.2458

Schmid, K., Reis-Filho, J. A., Harvey, E. S., & Giarrizzo, T. (2017). Baited remote underwater video as a promising nondestructive tool to assess fish assemblages in clearwater Amazonian rivers:

testing the effect of bait and habitat type. *Hydrobiologia*, *784*(1), 93–109. https://doi.org/10.1007/s10750-016-2860-1

Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, *9*(7), 671–675. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22930834

Schroeder, D. M., & Love, M. S. (2002). Recreational fishing and marine fish populations in California. *California Cooperative Oceanic Fisheries Investigations Reports*, *43*, 182–190.

ScimagoJR. (2017). Scimago Journal and Country Rank - Ethics Search Results [Web Page]. Retrieved August 18, 2017, from http://www.scimagojr.com/journalsearch.php?q=ethics

Scrapy. (2017). Scrapy | A Fast and Powerful Scraping and Web Crawling Framework [Web Page]. Retrieved March 3, 2017, from https://scrapy.org/

SDragon. (2006). SharpNLP 1.0.2529 Beta. [Computer Program]. Retrieved from sharpnlp.codeplex.com

Sea Angler Magazine. (2018). Sea Angler [Web Page]. Retrieved June 15, 2018, from https://www.seaangler.co.uk/

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*. Retrieved from http://www.statsmodels.org/dev/index.html

SEAFISH. (2016). *Overview of the Welsh sea bass fishing fleet*. Retrieved from http://www.seafish.org/media/1672991/overview_of_the_welsh_sea_bass_fishing_fleet_final_09-12-16_for_web.pdf

Shark Alliance. (2015). The Shark Alliance, UK shark tagging programme [Web Page]. Retrieved May 1, 2015, from http://www.ukshark.co.uk/

Shepperson, J., Murray, L. G., Cook, S., Whiteley, H., & Kaiser, M. J. (2014). Methodological considerations when using local knowledge to infer spatial patterns of resource exploitation in an Irish Sea fishery. *Biological Conservation*, *180*, 214–223. https://doi.org/10.1016/j.biocon.2014.10.013

Sherwood, G. D., & Grabowski, J. H. (2016). A comparison of cod life-history parameters inside and outside of four year-round groundfish closed areas in New England, USA. *ICES Journal of Marine Science*, *73*(2), 316–328. https://doi.org/10.1093/icesjms/fsv215

Shiffman, D. S., Macdonald, C., Ganz, H. Y., Hammerschlag, N., & Shi, D. S. (2017). Fishing practices and representations of shark conservation issues among users of a land-based shark angling online forum. *Fisheries Research*, *196*(December 2016), 13–26.

https://doi.org/10.1016/j.fishres.2017.07.031

Silver, J. J., & Campbell, L. M. (2005). Fisher participation in research: Dilemmas with the use of fisher knowledge. *Ocean and Coastal Management*, *48*(9), 721–741. https://doi.org/10.1016/j.ocecoaman.2005.06.003

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Silvertown, J., Harvey, M., Greenwood, R., Dodd, M., Rosewell, J., Rebelo, T., … McConway, K. (2015). Crowdsourcing the identification of organisms: A case-study of iSpot. *ZooKeys*, (480), 125–146. https://doi.org/10.3897/zookeys.480.8803

Siriwat, P., Nijman, V., Wildlife, O., & Sciences, S. (2018). Using online media-sourced seizure data to assess the illegal wildlife trade in Siamese rosewood. *Environmental Conservation*, 1–9. https://doi.org/10.1017/S037689291800005X

Smallwood, C. B., Pollock, K. H., Wise, B. S., Hall, N. G., & Gaughan, D. J. (2012). Expanding aerial-roving surveys to include counts of shore-based recreational fishers from remotely operated cameras: Benefits, limitations, and cost effectiveness. *North American Journal of Fisheries Management*, *32*(6), 1265–1276. https://doi.org/10.1080/02755947.2012.728181

Smith, B. D. (1999). A probabilistic analysis of decision-making about trip duration by Strait of Georgia sport anglers. *Canadian Journal of Fisheries and Aquatic Sciences*, *56*(6), 960–972. https://doi.org/10.1139/cjfas-56-6-960

Smith, C. M. (2005). Origin and uses of *primum non nocere*—above all, do no harm! *The Journal of Clinical Pharmacology*, *45*(4), 371–377.

Smith, N. S., & Zeller, D. (2015). Unreported catch and tourist demand on local fisheries of small island states: The case of the Bahamas, 1950-2010. *Fishery Bulletin*, *114*(1), 117–131. https://doi.org/10.7755/FB.114.1.10

Spampinato, C., Giordano, D., Salvo, R. Di, Fisher, R. B., & Nadarajan, G. (2010). Automatic Fish Classification for Underwater Species Behavior Understanding Categories and Subject Descriptors. In *Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams* (pp. 45–50). Firenze, Italy. https://doi.org/10.1145/1877868.1877881

St. Martin, K., & Hall-Arber, M. (2008). The missing layer: Geo-technologies, communities, and implications for marine spatial planning. *Marine Policy*, *32*(5), 779–786. https://doi.org/10.1016/j.marpol.2008.03.015

Stange, K. (2016). Building a knowledge base for management of a new fishery: Boarfish (Capros aper) in the Northeast Atlantic. *Fisheries Research*, *174*, 94–102. https://doi.org/http://dx.doi.org/10.1016/j.fishres.2015.08.023

Statista. (2018). Number of social media users worldwide from 2010 to 2021 [Web Page]. Retrieved August 10, 2018, from https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/

Steffe, A. S., Murphy, J. J., & Reid, D. D. (2008). Supplemented access point sampling designs: A cost-effective way of improving the accuracy and precision of fishing effort and harvest estimates derived from recreational fishing surveys. *North American Journal of Fisheries Management*, *28*(4), 1001–1008. https://doi.org/10.1577/m06-248.1

Steffe, A. S., Taylor, S. M., Blight, S. J., Ryan, K. L., Desfosses, C. J., Tate, A. C., … Wise, B. S. (2017). *Framework for integration of data from remotely operated cameras into recreational fishery assessments in Western Australia* [Report]. Retrieved from http://fish.wa.gov.au/Documents/research_reports/frr286.pdf

Stephenson, R. L., Paul, S., Pastoors, M. A., Kraan, M., Holm, P., Wiber, M., … Benson, A. (2016). Integrating fishers' knowledge research in science and management. *ICES Journal of Marine Science*, *73*(6), 1459–1465. https://doi.org/10.1093/icesjms/fsw025

Stevens, T. (1996). S.39 - 104th Congress (1995-1996): Sustainable Fisheries Act. Retrieved from https://www.congress.gov/bill/104th-congress/senate-bill/39

Stolk, P. (2009). *Angling participation: interim report part 1 – the social and community benefits of angling research* [Report]. Substance. Retrieved from www.resources.anglingresearch.org.uk

Stowell, D. (2018). Warblr, The Birdsong Recognition App [Web Page]. Retrieved August 1, 2018, from https://www.warblr.co.uk/warblrteam/

Strachan, N. J. C. (1993). Length measurement of fish by computer vision. *Computers and Electronics in Agriculture*, *8*(2), 93–104. https://doi.org/10.1016/0168-1699(93)90009-P

Struthers, D. P., Danylchuk, A. J., Wilson, A. D. M., & Cooke, S. J. (2015). Action cameras: Bringing aquatic and fisheries research into view. *Fisheries*, *40*(10), 502–512. https://doi.org/10.1080/03632415.2015.1082472

Sun, X., Shi, J., Dong, J., & Wang, X. (2016). Fish Recognition from Low-resolution Underwater Images. In *2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics* (pp. 471–476). Datong, China. https://doi.org/10.1109/CISP-BMEI.2016.7852757

Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional network cascade for facial point detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition, 3476–3483. https://doi.org/10.1109/CVPR.2013.446

Svensson, G. (2016). Do you have any particular favorite place?: Hunters' and anglers' secrets meet tourism in northern Norway. *Arctic Anthropology*, *53*(1), 58–68.

Szeliski, R. (2010). *Computer Vision: algorithms and applications*. [Book], Springer Science & Business Media. https://doi.org/10.1007/978-1-84882-935-0

Tamou, A. Ben, Benzinou, A., Nasreddine, K., & Ballihi, L. (2018). Underwater Live Fish Recognition by Deep Learning. In A. Mansouri, A. El Moataz, F. Nouboud, & D. Mammass (Eds.), *Image and Signal Processing. ICISP 2018. Lecture Notes in Computer Science, vol 10884* (pp. 275–283). Springer, Cham. https://doi.org/10.1007/978-3-319-94211-7_30

Tarrant, M. A., & Manfredo, M. J. (1993). Digit preference, recall bias, and nonresponse bias in self reports of angling participation. *Leisure Sciences*, *15*(3), 231–238. https://doi.org/10.1080/01490409309513202

The White House. (2010, July). Executive Order 13547 - Stewardship of the Ocean, Our Coasts, and the Great Lakes. Retrieved February 10, 2018, from https://www.whitehouse.gov/the-press-office/executive-order-stewardship-ocean-our-coasts-and-great-lakes

Tillett, R., McFarlane, N., & Lines, J. (2000). Estimating dimensions of free-swimming fish using 3D point distribution models. *Computer Vision and Image Understanding*, *79*(1), 123–141. https://doi.org/10.1006/CVIU.2000.0847

TNS Global. (2014a). *The Great Britain Day Visit Survey (2014). GB Day Visits 2013. On behalf of Visit England, Visit Scotland and Visit Wales.* [Report]. Visit England;Visit Wales;Visit Scotland. Retrieved from https://data.gov.uk/dataset/eea42780-2474-49b8-a646-486e5f2f95ba/great-britain-day-visits-survey

TNS Global. (2014b). *The Great Britain Tourism Survey. Volumes and Values of Domestic Tourism in Great Britain – 2013. On behalf of Visit England, Visit Scotland and Visit Wales.* [Report]. Visit England;Visit Wales;Visit Scotland. Retrieved from https://www.visitbritain.org/sites/default/files/vb-corporate/Documents-Library/documents/England-documents/gb_tourist_2013_v2.pdf

Townsend, L., & Wallace, C. (2016). *Social Media Research: A Guide to Ethics* [Report]. Aberdeen. Retrieved from https://www.gla.ac.uk/media/media_487729_en.pdf

Tsikerdekis, M., & Zeadally, S. (2014). Online deception in social media. *Communications of the ACM*, *57*(9), 72–80. https://doi.org/10.1145/2629612

UK Parliament. Animals (Scientific Procedures) Act 1986 (1986). Statute Law Database. Retrieved from http://www.legislation.gov.uk/ukpga/1986/14/contents

UK Parliament. Marine and Coastal Access Act 2009 (2009). United Kingdom. Retrieved from http://www.legislation.gov.uk/ukpga/2009/23/contents

United Kingdom Hydrographic Office. (2013). UKHO British Admiralty Maps, published under license by Digimap [Web Page]. Retrieved December 20, 2017, from http://digimap.edina.ac.uk/digimap/home

United States Code. Fishery Conservation and Management. 16 U.S.C. Ch. 38, Pub. L. No. 16 (1976). U.S.C. Retrieved from http://uscode.house.gov/view.xhtml?path=/prelim@title16/chapter38&edition=prelim

United States Code. Limitations on exclusive rights: Fair use. 17 U.S.C. § 107 (2011), Pub. L. No. 17 (2011). U.S.C.

van der Hammen, T., & de Graaf, M. (2015). *Recreational fisheries in the Netherlands: analyses of the 2012-2013 online logbook survey, 2013 online screening survey and 2013 random digit dialing screening survey* [Report] (Vol. C042/15). IMARES. Retrieved from http://library.wur.nl/WebQuery/titel/2088516

van der Hammen, T., de Graaf, M., & Lyle, J. M. (2016). Estimating catches of marine and freshwater recreational fisheries in the Netherlands using an online panel survey. *ICES Journal of Marine Science*, *73*(2), 441–450. https://doi.org/10.1093/icesjms/fsv190

van der Walt, S., Colbert, C. S., & Varoquaux, G. (2011). The NumPy Array: A structure for efficient numerical computation. *Computing in Science & Engineering*, *13*, 22–30. https://doi.org/10.1109/MCSE.2011.37

van Helmond, A. T. M., Chen, C., & Poos, J. J. (2017). Using electronic monitoring to record catches of sole (*Solea solea*) in a bottom trawl fishery. *ICES Journal of Marine Science*, *74*(5), 1421–1427. https://doi.org/10.1093/icesjms/fsw241

van Mierlo, T. (2014). The 1% rule in four digital health social networks: an observational study. *Journal of Medical Internet Research*, *16*(2).

van Poorten, B. T., Carruthers, T. R., Ward, H. G. M., & Varkey, D. A. (2015). Imputing recreational angling effort from time-lapse cameras using an hierarchical Bayesian model. *Fisheries Research*, *172*, 265–273. https://doi.org/10.1016/j.fishres.2015.07.032

van Voorhees, D. A., Sminkey, T. R., Schlechte, J. W., Donaldson, D. M., Anson, K. J., O'Hop, J. R., … Zales, R. F. I. I. (2002). *The New Marine Recreational Fishery Statistics Survey Method for Estimating Charter Boat Fishing Effort*. (R. L. Creswell, Ed.), *Proceedings of the Fifty-Third Annual Gulf and Caribbean Fisheries Institute*. [Book], Gulf and Caribbean Fisheries Institute, c/o Harbor Branch Oceanographic Institution, Inc. 5600 US 1 North Fort Pierce FL 34946 USA, [URL:http://www.gcfi.org]. Retrieved from http://search.proquest.com/docview/18611443?accountid=14874

Vandaele, R., Aceto, J., Muller, M., Péronnet, F., Debat, V., Wang, C. W., … Marée, R. (2018). Landmark detection in 2D bioimages for geometric morphometrics: A multi-resolution tree-based approach. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-017-18993-5

Vázquez, F. J. S., & Muñoz-Cueto, J. A. (2014). *Biology of European sea bass*. [Book], CRC Press.

Veiga, P., Ribeiro, J., Gonçalves, J. M. S., & Erzini, K. (2010). Quantifying recreational shore angling catch and harvest in southern Portugal (north-east Atlantic Ocean): implications for conservation and integrated fisheries management. *Journal of Fish Biology*, *76*(9), 2216–2237. https://doi.org/10.1111/j.1095-8649.2010.02665.x

Venturelli, P. A., Hyder, K., & Skov, C. (2017). Angler apps as a source of recreational fisheries data : opportunities , challenges and proposed standards. *Fish and Fisheries*, *18*, 578–595. https://doi.org/10.1111/faf.12189

Vince, J. (2014). Oceans governance and marine spatial planning in Australia. *Australian Journal of Maritime & Ocean Affairs*, *6*(1), 5–17. https://doi.org/10.1080/18366503.2014.888137

Vølstad, J. H., Pollock, K. H., & Richkus, W. A. (2006). Comparing and combining effort and catch estimates from aerial–access designs as applied to a large-scale angler survey in the Delaware River. *North American Journal of Fisheries Management*, *26*(3), 727–741. https://doi.org/10.1577/M04-146.1

W3C. (2017). World Wide Web Consortium (W3C) home page [Web Page]. Retrieved March 2, 2017, from https://www.w3.org/

Walther, J. B. (2002). Research ethics in Internet-enabled research: human subjects issues and methodological myopia. *Ethics and Information Technology*, *4*(4), 205–216. https://doi.org/10.1023/A:1021368426115

Wang, E. S.-T., & Chen, L. S.-L. (2012). Forming relationship commitments to online communities: The role of social motivations. *Computers in Human Behavior*, *28*(2), 570–575. https://doi.org/10.1016/J.CHB.2011.11.002

Weinstein, B. G. (2015). MotionMeerkat: Integrating motion video detection and ecological monitoring. *Methods in Ecology and Evolution*, *6*(3), 357–362. https://doi.org/10.1111/2041-210X.12320

Wells, R. S., Hofmann, S., & Moors, T. L. (1998). Entanglement and mortality of bottlenose dolphins, *Tursiops runcatus*, in recreational fishing gear in Florida. *Fishery Bulletin*, *96*(3), 647–650. Retrieved from https://spo.nmfs.noaa.gov/content/entanglement-and-mortality-bottlenose-dolphins-tursiops-runcatus-recreational-fishing-gear

Welsh Government. (2017). Draft Welsh National Marine Plan. WGS31640 [Web Page]. Retrieved June 1, 2018, from https://beta.gov.wales/draft-welsh-national-marine-plan

Weltersbach, M. S., & Strehlow, H. V. (2013). Dead or alive—estimating post-release mortality of Atlantic cod in the recreational fishery. *ICES Journal of Marine Science: Journal Du Conseil*, *70*(4), 864–872. https://doi.org/10.1093/icesjms/fst038

Westera, M., Lavery, P., & Hyndes, G. (2003). Differences in recreationally targeted fishes between protected and fished areas of a coral reef marine park. *Journal of Experimental Marine Biology and Ecology*, *294*(2), 145–168. https://doi.org/10.1016/S0022-0981(03)00268-5

White, D. J., Svellingen, C., & Strachan, N. J. C. (2006). Automated measurement of species and length of fish by computer vision. *Fisheries Research*, *80*(2–3), 203–210. https://doi.org/10.1016/j.fishres.2006.04.009

Wickham, H. (2016). rvest: Easily Harvest (Scrape) Web Pages [Web Page]. Retrieved February 10, 2017, from https://cran.r-project.org/package=rvest

Wilde, G. R., & Pope, K. L. (2013). Worldwide trends in fishing interest indicated by internet search volume. *Fisheries Management and Ecology*, *20*(2–3), 211–222. https://doi.org/10.1111/fme.12009

Willemen, L., Cottam, A. J., Drakou, E. G., & Burgess, N. D. (2015). Using social media to measure the contribution of red list species to the nature-based tourism potential of African protected areas. *PLOS ONE*, *10*(6), e0129785. https://doi.org/10.1371/journal.pone.0129785

Williams, S. M., Holmes, B. J., & Pepperell, J. G. (2015). The novel application of non-lethal citizen science tissue sampling in recreational fisheries. *PLOS ONE*, *10*(9), e0135743. https://doi.org/10.1371/journal.pone.0135743

Wilson, C., & Tisdell, C. (2003). Conservation and economic benefits of wildlife-based marine tourism: Sea turtles and whales as case studies. *Human Dimensions of Wildlife*, *8*(1), 49–58. https://doi.org/10.1080/10871200390180145

Wise, B. S., & Fletcher, W. J. (2013). *Determination and development of cost effective techniques to monitor recreational catch and effort in Western Australia demersal finfish fisheries. Final Report for FRDC Project 2005/034 and WAMSI Subproject 4.4.3. Fisheries Report No. 245* [Report]. Retrieved from http://www.fish.wa.gov.au/Documents/research_reports/frr245.pdf

Wong, S. C., Gatt, A., Stamatescu, V., & McDonnell, M. D. (2016). Understanding Data Augmentation for Classification: When to Warp? In *2016 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2016*.

https://doi.org/10.1109/DICTA.2016.7797091

Wood, S. a, Guerry, A. D., Silver, J. M., & Lacayo, M. (2013). Using social media to quantify nature-based tourism and recreation. *Scientific Reports*, *3*, 2976. https://doi.org/10.1038/srep02976

World Health Organisation. (2015). WHO | Ethical standards and procedures for research with human beings [Web Page]. Retrieved July 16, 2017, from http://www.who.int/ethics/research/en/

World Intellectual Property Organization. (1979). WIPO-Administered Treaties: Berne Convention for the Protection of Literary and Artistic Works. Retrieved July 18, 2017, from http://www.wipo.int/treaties/en/text.jsp?file_id=283698

World Medical Association. (2001). Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, *79*(4), 373.

Wynne, B. (1992). Misunderstood misunderstanding: Social identities and public uptake of science. *Public Understanding of Science*, *1*(3), 281–304.

Yates, D., & Paquette, S. (2010). Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. In *Proceedings of the 73rd ASIS&T Annual Meeting on Navigating Streams in an Information Ecosystem - Volume 47* (p. 42). American Society for Information Science. Retrieved from https://dl.acm.org/citation.cfm?id=1920391

Yates, K. L., & Schoeman, D. S. (2013). Spatial access priority mapping (SAPM) with fishers: A quantitative GIS method for participatory planning. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068424

Yeo, I.-K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, *87*(4), 954–959.

Zaiontz, C. (2017). Real Statistics Using Excel [Web Page]. Retrieved December 17, 2017, from http://www.real-statistics.com/free-download/real-statistics-resource-pack/

Zeller, D., Booth, S., Davis, G., & Pauly, D. (2007). Re-estimation of small-scale fishery catches for U.S. flag-associated island areas in the western Pacific: The last 50 years. *Fishery Bulletin*, *105*(2), 266–277.

Zeller, D., Rossing, P., Harper, S., Persson, L., Booth, S., & Pauly, D. (2011). The Baltic Sea: Estimates of total fisheries removals 1950-2007. *Fisheries Research*, *108*(2–3), 356–363. https://doi.org/10.1016/j.fishres.2010.10.024

Zhang, G. (2017). Smartphones now account for half the world's mobile connections [Web Page]. Retrieved December 11, 2017, from https://www.gsmaintelligence.com/research/2017/02/smartphones-now-account-for-half-the-worlds-mobile-connections

Zhang, J., Kan, M., Shan, S., & Chen, X. (2016). Occlusion-Free Face Alignment: Deep Regression Networks Coupled with De-Corrupt AutoEncoders. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3428–3437). IEEE. https://doi.org/10.1109/CVPR.2016.373

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(11), 1330–1334. https://doi.org/10.1109/34.888718

Zhou, B., Pei, J., & Luk, W. (2008). A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor. Newsl.*, *10*(2), 12–22. https://doi.org/10.1145/1540276.1540279

Zion, B., Alchanatis, V., Ostrovsky, V., Barki, A., & Karplus, I. (2007). Real-time underwater sorting of edible fish species. *Computers and Electronics in Agriculture*, *56*(1), 34–45. https://doi.org/10.1016/j.compag.2006.12.007

Zion, B., Shklyar, A., & Karplus, I. (1999). Sorting fish by computer vision. *Computers and Electronics in Agriculture*, *23*(3), 175–187. https://doi.org/10.1016/S0168-1699(99)00030-7

Zion, B., Shklyar, A., & Karplus, I. (2000). In-vivo fish sorting by computer vision. *Aquacultural Engineering*, *22*(3), 165–179. https://doi.org/10.1016/S0144-8609(99)00037-0

Zooniverse. (2017). Zooniverse: The list of active projects [Web Page]. Retrieved February 10, 2017, from https://www.zooniverse.org/projects?status=live

Zoph, B., & Le, Q. V. (2017). Neural Architecture Search with Reinforcement Learning. In *International Conference on Learning Representations*. Toulon, France. Retrieved from http://arxiv.org/abs/1611.01578